A framework for tiered informed consent for health genomic research in Africa

A generic framework for providing participant information and implementing a tiered consent process for health genomic research in Africa can help to harness global health benefits from sharing and meta-analysis of African genomic data while simultaneously respecting and upholding the autonomy and individual choices of African research participants.

Victoria Nembaware, Katherine Johnston, Alpha A. Diallo, Maritha J. Kotze, Alice Matimba, Keymanthri Moodley, Godfrey B. Tangwa, Rispah Torrorey-Sawe and Nicki Tiffin

African human genome research is advancing rapidly, owing to falling sequencing costs and international interest in African genomic data: the diversity of African genomes can provide novel insights into biological and etiological mechanisms, thereby promoting diagnostic, prognostic and therapeutic advances for populations in Africa and the rest of the world[1,2]. Conducting genomic research in Africa can be logistically challenging[3,4], but equally challenging is recruiting African participants—many of whom have knowledge- and/or poverty-related vulnerabilities[5,6]—while ensuring that they are properly informed and truly consenting, and that they retain their autonomy and agency (examples in refs. [5,7–12]).

Dynamic consent models, such as those using ongoing engagement through online media[13,14], ensure autonomy and choice for participants. However, these models cannot be implemented in many African environments, owing to suboptimal internet and smartphone access, and poor digital literacy. In broad-consent models, participants consent to all future use of their samples under the oversight of an access committee, but this consent comes at the cost of the autonomy and individual preferences of participants[10,15]. Tiered informed consent addresses these challenges by providing detailed information about the intended specimen/data use and storage, thus enabling participants to individually select a level of specimen and/or data sharing through responses to specific questions[10,16–18].

The launch of the H3Africa genotyping chip, with 2.7 million African-specific genomic variants[19], has increased opportunities for meta-analyses combining multiple cohorts of African participants. To undertake such studies with an ethical mandate, research participants must be properly informed, and their data/sample-use preferences must be accurately understood, faithfully recorded and implemented with integrity. We propose a framework for undertaking ethically sound, tiered-informed-consent processes in Africa, which provides a comprehensive guide on compiling participant information and an informed-consent template for capturing each participant's consent information and mapping it to data-use ontologies.

Methods

The framework is derived from the authors' combined experiences with informed consent in Africa. A generic template for participant information (Supplementary Note) is provided, and validated translation into the participants' primary language(s) is recommended. Framework tenets include ease of understanding for field-workers and participants, and practical administration in a busy facility (such as clinics). We describe the core components of tiered consent for competent adult participants, although not all elements may be appropriate for every study. We provide recommendations for data capture and standardization of participant consent information.

What information should be provided in the informed-consent documents?

Ethical research requires balancing benefits and risks at the micro level for individuals, the meso level for communities and the macro level for populations. Although ethics review boards consider all levels in assessing a research study, an individual's right to decline or accept participation remains paramount, and participants must be provided with necessary and sufficient information to support this right. We present core concepts and highlight how localized knowledge can be incorporated in providing information to participants.

Information about genetics. African colloquialisms often speak to an inherent understanding of heredity. For example, the Shona proverb "Mhembwe rudzi inozvara mwana ane kazhumu" translates to "The child of a duiker [small antelope] is a duiker," and its equivalent in English is "Like father, like son." Such local expressions can be harnessed in explaining heredity and genetic concepts by using anecdotally accessible, emotionally neutral examples such as height or facial similarity. Caucasian-centric examples, such as eye or hair color, are often inappropriate for African participants.

Information about genetics and health can be related to locally prevalent health conditions, while always providing clarity about complex risk factors to prevent misunderstanding, anxiety or family conflict—for example, an explanation that genetic factors might influence susceptibility to malaria, but environmental factors are the chief drivers of becoming infected.

The focus of the study, and who is doing the research. Clear, simple language and local names for health conditions can simplify explanation of the research question. For example, "We want to understand whether genes affect how likely someone is to get sick from bilharzia" is more accessible than "The primary study objective is to elucidate the genetic etiology of schistosomiasis" for an African population exposed to schistosomes. However, information accuracy supports transparency, and researchers should not confuse straightforward language with incomplete information. During recruitment of controls, researchers can explain that comparing samples from people with and without an illness can help understand what contributes to getting the illness. Local researchers and institutions should always be named as the primary contact, to ensure that they are familiar, un-intimidating, identifiable and contactable for participants.

What will you be asked to provide or do in this study?. A brief, simple explanation of exactly what will be requested in terms of visits, data and sample collection, and common-use estimates of collection volumes can inform participants about the collection of blood or saliva. For example, referring to "about two teaspoons" of blood is more readily understood by participants than "10 ml."

What are the potential risks and benefits of this study?. There are often no direct benefits to participants in genome research; in such cases, the lack of benefits should be honestly stated. However, researchers should also recognize and respect that African participants may also value altruistic behavior and contributing to the well-being of others or the advancement of science, regardless of their circumstances[20,21]. Research studies may be misunderstood by participants as an offer of additional health care or an opportunity for cure, and if such benefits will not be provided, this must be explicitly stated.

Communicating risks is complicated by the unknowable nature of future risks associated with genomic data: the rapidly evolving landscape makes cataloging future possible uses impossible. Existing risks include re-identification of individuals and exposure of personal health information in the event of data breach or inappropriate data reuse[22–24], as well as stigmatization of families, communities or ancestral groups. Another risk to be communicated to participants is discovery of information about the participant's health or information that might negatively affect family members or local community members who did not necessarily consent to the study. Describing the risks alongside clear, practical plans for risk mitigation can reassure participants that risks have been appropriately identified and planned for.

Remuneration for costs incurred by study participation, or refreshments provided, should be detailed separately to avoid confusion with 'study benefits'.

Privacy protection, and data- and specimen-protection protocols. The processes and infrastructure in place to protect data and specimens should be briefly outlined to assure participants that privacy protection is in place. Sample storage locations and security measures should be detailed—including the geographical locations of storage and use, what data will be generated from the samples, and how and by whom those data will be used further. Information can be provided about the committees that will oversee

access and the plans for sharing aggregated or group-level data with collaborators, scientific journals, international collaborators or online platforms.

Return of results. General study findings can be communicated to participants in a context-appropriate process involving posters and pamphlets distributed in recruitment facilities, cell-phone messaging or website updates. Community meetings with the researchers to provide feedback on findings should occur only if participant confidentiality can be protected, especially for studies on sensitive health issues. The return of findings from genome-related research requires approval by a research ethics committee and medical specialists who can determine whether the findings (primary or incidental) meet criteria for return to individual participants and are actionable. The consent process should support the participants' right not to know certain results, whether actionable or not actionable, within their current context. Furthermore, the process of disclosing research results should involve professionals with the appropriate expertise, and a summary of this planned process should be communicated to participants.

For the initial study for which participants are recruited, participant information should detail whether individual results will be made available; if so, the process for returning results should be clearly communicated. For example, participant information might state that no results will be returned to individuals because the findings will not be sufficient to provide accurate health-related information; that a doctor will provide results during a clinical consultation with the participant; or that an individual report will be provided by appropriately registered medical scientists. This section must also inform participants of what action will be taken if a communicable disease is identified, including how the participant will be informed, plans for linkage to care and the process for infectious diseases that must legally be reported to a central/national registry.

Information should describe how study findings will be shared with the participant community, for example through a project website (with provided URLs) and/or newsletters by e-mail or hard copy. The intended publication of results within the research sector can also be described.

Who can be contacted with questions or concerns, and how to withdraw consent. Names and contact information must be provided for participant questions or concerns. In addition, contact details for

an oversight body independent of the study researchers, such as an ethics review board, which can address concerns impartially, should be provided. Clear instructions on an easy process to withdraw from the study must be provided, with reassurance that withdrawal will not affect access to standard health care if recruitment takes place in a health facility.

Questions to be asked for each component of the tiered consent

We propose a series of questions that each define a tier of the consent process. The first question defines inclusion in the initial study, and a response of 'yes' is required to proceed. The questions that then follow are designed to be freestanding, and each can be independently agreed to or declined, so that participants define a particular combination of data/biospecimen uses with which they feel comfortable.

Question 1: Agreement for collection of data/biospecimens for the primary study. "Do you agree for us to collect this saliva/blood sample and your health information for this study that we have described on how genes might affect [specific health phenotype]?"

This question defines consent to participate in the current (primary) study for which participants are being recruited. Often in health research, primary studies examine a specific disease, for example, tuberculosis or hypertension.

Question 2: Agreement for secondary data/biospecimen use for other studies on the same phenotype/health condition. "Do you agree for us to use your genetic sample together with your health information in other studies in the future on the effects of genes on [specific health phenotype]?"

This question provides an opportunity for individuals to contribute to furthering research on a specific disease that they might feel particularly strongly about because of their personal experience, without committing to general research use in other types of health research.

Question 3: Agreement for general secondary data/biospecimen use in other unrelated studies. "Do you agree for us to use your genetic sample together with your health information for other studies in the future to study the effects of genes on *other* conditions or biological processes?"

This option enables general re-use of data and samples in future studies that are not yet defined or known, while providing the option for those who are not comfortable with wider re-use of their data/sample to clearly define

a boundary of health or biological function research for secondary use.

Question 4: Agreement for inclusion in aggregated data (for example, genome summary data) for the study. "Sometimes researchers combine the genetic information from everyone in the study and provide a summary of genetic data for the whole group. Do you agree for us to use your information when providing combined information about the whole research group (*x* total individuals in this study)?"

A recent policy change by the US National Institutes of Health[25] about open sharing of genomic summary data from studies has prompted this consideration. The risk of stigmatization or of discrimination against ancestral groups is substantial in Africa, where the genetic distance between subgroups can be large, and communities can be small and easily identified. Both historically and recently, the term 'ethnicity' has been a sensitive or volatile identifier. We believe, therefore, that consent should be obtained before individuals' data are released within an aggregate dataset. Providing the total number of participants can give an indication of the likelihood of being re-identified.

Question 5: Agreement for re-contact for follow-on studies. "Sometimes, what we find from a study like this might lead to new studies being done in the future. Can other researchers contact you in the future to invite you to take part in other research studies?"

Re-contact of individuals can be difficult, owing to the high geographic mobility of some African populations, shared or transitory cell-phone ownership, informal residential addresses and limited modes of contact. Re-contact outside the consented study violates the privacy of participants and constitutes inappropriate or even illegal secondary use of personal information unless consent has been specifically given for such re-contact. However, some individuals may agree to be contacted about other studies, particularly to facilitate access to specialist health care.

Question 6: Agreement for return of defined genetic findings from the current study. "In this study, we hope to identify genetic factors that mean that someone is more likely to [have an outcome, such as susceptibility to a disease]. If someone has this genetic factor, there is [no treatment/we recommend treatment *x*]. If we find during this study that you have this kind of genetic

factor, would you like us to tell you this information?"

When genomic studies investigate specific genetic factors affecting health, consent for return of results is easier to delineate. For non-African settings, recommendations suggest that participants be asked whether they wish to receive clearly identified, actionable and/or non-actionable genetic findings from the study[26]. The practical implications, including treatment recommendations, can be defined in advance for the participants, who can use contextualized information and personal preferences to decide which results they wish to receive. Challenges arise around limited health care access and whether participants can access or afford a recommended intervention. Researchers and ethics review boards must address these questions within the study context to maximize benefits through diagnostic and therapeutic information while minimizing emotional and societal harm in cases in which such information cannot improve health outcomes.

Question 7: Agreement for return of unanticipated genetic findings from future studies. "Sometimes what we find from our research might include new information about your health. Would you like us to contact you again if we believe we have new information that may directly affect your health—if there is some kind of action or treatment that might be able to help you with the health issue (yes/no) and if there is *no* kind of action or treatment that might be able to help you with the health issue (yes/no)?"

Implications for returning results become more nuanced and difficult to communicate for secondary use of data/biospecimens, because the nature of future genetic or health findings is currently unknown. We propose that a distinction should be made between actionable and non-actionable results, to accommodate individuals who prefer to know nothing about their etiological genetic background; those who wish to know only information with available interventions; or those who wish to know all genetic components to their health, both actionable or unactionable. The principal investigators, data-access committee and ethics review board must ensure that secondary-analysis protocols include plans for the return of results and assessments of appropriate available interventions in a consultative, informed and supported manner.

How to store informed-consent choices
Common practice in Africa is to capture participant consent in hard copy, without sharing the consent with secondary users.

Digitalization of informed consent usually consists of an e-consent signature or scanned copy of the signed consent form in a non-machine-readable format. These practices do not support tiered consent, which captures multiple accessible, query-able and actionable choices for individuals. Consent interpretation and automated selection of data/samples for secondary use can also be challenging, owing to unstandardized and heterogeneous consent questions; moreover, different consent standards between fields, such as clinical and genomic research, also limit cross-disciplinary data sharing[27]. Ontologies such as the Informed Consent Ontology (ICO)[28] and Data Usage Ontology (DUO)[27,29] provide standardized terminology and systems to semantically label samples and data with their usage restrictions: ICO captures the process of obtaining informed consent, whereas DUO describes consent and governance categories, coding 19 primary and secondary data-use cases. Tools, templates and software can assist with practical implementation of DUO[30] and capture metadata of consent information, restrictions and requirements for a study, but they do not capture individual consent choices[30]. DUO[27] lacks some key consent codes for return of actionable and nonactionable, anticipated and unanticipated findings, aggregated data use and consent for participant re-contact after the study. However, consent ontologies and coding are likely to evolve to meet such requirements.

Simple, low-cost, systematic strategies are needed to capture, store, share and take action on individuals' tiered consent choices through commonly used platforms (for example REDCap[31]). Each informed-consent question requires binary response options as checkbox items; the case-report form should be versioned; and any consent changes should be exported to relevant laboratories, biobanks, data repositories or data analysts for action to be taken if sample destruction or data deletion is required. An example of data-use consent information and corresponding DUO codes for a sickle-cell disease genetic study is shown in Box 1.

Each tier of consent is coded as a binary variable (yes/no, 0/1), with the date of consent. When individuals modify their consent information, a database structure can allow for the addition of new consent data with a combined, unique date-study ID key. A 'current' flag facilitates identification of most recent consent data (Table 1). The binary matrix design enables simple, intuitive data capture of consent information, and combinations of binary values are subsequently mapped to consent codes from DUO or other ontologies as they

Box 1 | Illustration of the coding of tiered-consent forms for two participants being recruited into a sickle-cell disease genetic study

Participant 1

1.1. Do you agree for us to use your genetic sample together with your health information for this study on the effects of genes on sickle-cell disease?
Answer: Yes
1.2. Do you agree for us to use your genetic sample together with your health information for other studies in the future on the effects of genes on sickle-cell disease? Answer: Yes
1.3. Do you agree for us to use your genetic sample together with your health information for other studies in the future to study the effects of genes on other conditions or biological processes?
Answer: No
DUO requirements/restriction description: Disease-specific research and clinical care
DUO requirement code: DS-(XX)(CC)

Participant 2

1.1. Do you agree for us to use your genetic sample together with your health information for this study on the effects of genes on sickle-cell disease? Answer: Yes
1.2: Do you agree for us to use your genetic sample together with your health information for other studies in the future on the effects of genes on sickle-cell disease? Answer: Yes
1.3: Do you agree for us to use your genetic sample together with your health information for other studies in the future to study the effects of genes on other conditions or biological processes? Answer: Yes
DUO requirements/restriction description: Use of the data limited to health/medical/biomedical research but *not* population origin/ancestry
DUO requirement code: HMB(CC)

become available or are updated. Future work to improve data-capture fidelity may include image processing to automatically capture and code responses from signed consent documents. A consent 'metadata' record can also describe the type of informed-consent questions asked during that study.

Discussion

This framework is a practical guide for preparing informed-consent documents for consultation with people considering participation in a genetic or genomic research study, assuming that informational tools including pamphlets, videos and flip charts are also used, and a community-engagement process precedes and continues during the study. Although beyond the scope of this publication, community engagement before human genomic research in Africa is paramount[15,32].

Individual informed consent addresses the micro level; however, at the meso and macro levels, families, communities, religious groups, ancestral groups and populations are all affected by genomic data from individuals[32,33], particularly in Africa, where populations are highly diverse and often genetically unique and re-identifiable, and ancestry-based perceptions have previously fueled life-threatening discord. Consent processes must by necessity be situated in this broader context[34]. Sufficient time should be given between providing information to potential participants and the enrolment

visit, to respect family and communal decision-making; in addition, providing a recap of consent options in multiple visits is advisable.

While compiling the framework, we identified issues requiring further exploration regarding the recommendations for implementation of tiered informed consent in Africa. These include the following.

Return of results. Incidental or unforeseen results from secondary data analysis are challenging to address during tiered informed consent and must be considered within the context of each study. This area has a high risk of unforeseeable harm to participants, families and communities, and ongoing research into the return of results and how to define 'actionable results' in Africa and low- and middle-income countries (LMICs) in general is essential.

Informed consent during times of crisis. The recent Ebola crisis in West Africa highlighted issues in using biological samples toward the common good and in informed-consent processes in times of crisis[35–37]. We must explore waiver of informed consent in times of crisis in which minimum use of samples/data for public good might override individual rights, or individuals might be too ill to consent: within the tiered-consent model, levels of consent that might be acceptably waived and those that should remain non-negotiable may be identified.

Consent for vulnerable populations. We reiterate that these consent guidelines are intended for competent and autonomous adult participants. Particularly in LMICs and in Africa, many participants may be vulnerable, including those with limited access to health care and socioeconomic resources, children and adolescents, disenfranchised women, persecuted ancestral groups and people marginalized or criminalized because of their sexual orientation and/or gender identity. For studies involving potentially vulnerable participants, we recommend that specialist advice be sought to ensure an appropriate informed-consent process.

Intended commercialization of study findings. Intended commercialization of findings provides a special case for consent, and benefit-sharing agreements with participants require research to address the potential for coercion or inducement into participation with promise of financial rewards, community consultation around appropriate avenues to return benefits, and social constructs and community pressure to participate that might arise from promises of community-level benefits with sufficient participation rates.

Legislative constraints for consent processes. Protection of participant data and confidentiality is increasingly protected by new legislation, and consent processes must comply with national laws and regulations. In South Africa, for example, the Protection of Personal Information Act may affect whether broad consent can be legally obtained from participants. Local legal advice is essential to ensuring that local legislation is respected.

Consent to studies of population origins. We have intentionally omitted secondary-use consent for studies of ancestry or population origins in this framework, because of the complexity of risks and future-use cases for such research. Recruiting health-research participants entails approaching ill people who may have personal motivations to participate in health research. We propose that population-diversity genome-research consent should be addressed separately to avoid a 'bait-and-switch' approach in which vulnerable participants consent to health research but provide consent to population-diversity research as a secondary mechanism without necessarily understanding the full risks or implications of such research.

We present this framework as a starting point for implementing tiered informed consent in Africa, providing a generic

Table 1 | Example of data coding and capture for tiered-consent questions, with retrospective mapping to DUO

	Captured and stored binary variables from the informed-consent process, showing capture of response to each tier											Retrospective mapping to current ontologies; can be re-mapped as ontologies are updated or replaced[c]	
Study ID	Date of consent	Version consent document	Is current	Use for primary study of health condition	Use for other studies on this condition	Use for other studies on health or general biology	Use of aggregate data for entire study	Re-contact for future studies	Return of specific, described genetic results from primary study	Return of undefined, actionable results from other studies	Return of undefined, non-actionable results from other studies	DUO consent description	DUO code
ID_1	2010-10-20	1	1	0	NA	NA	NA	NA	NA	NA	NA	No consent given	
ID_2	2010-10-20	1	0	1	1	0	0[a]	1	1	1	1	Use of data must be related to [disease]	DS-[XX](CC)
ID_2[b]	2010-10-25	1	1	1	0	0	0	0	1	0	0	Use of data is limited to use within an approved project	PS
ID_3	2010-10-20	1	1	1	1	0	1	1	1	1	1	Use of data must be related to [disease]	DS-[XX](CC)
ID_4	2010-10-20	1	1	1	1	1	1	0	1	1	0	Use of data is limited to health/medical/biomedical research but *not* population origin/ancestry	HMB(CC)

[a] Currently, aggregate data use, re-contact and return of results are not coded in DUO. [b] Shows change in consent for individual ID_2 and use of the 'current' 0/1 flag. [c] Fields are retrospectively mapped to DUO by using stored variables and can be updated. NA, not applicable.

example of participant information and tiered-consent questions to be adapted for individual contexts in Africa, other LMICs and beyond. This example is intended as an illustration of ways to address different scenarios for participant information and consent questions. We welcome dialog and recommendations for improvements to this framework to benefit African participants in the future.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article. ⌐

Victoria Nembaware[1,13], Katherine Johnston[2,13], Alpha A. Diallo[3,4], Maritha J. Kotze[5,6], Alice Matimba[7], Keymanthri Moodley[8], Godfrey B. Tangwa[9], Rispah Torrorey-Sawe[5,10] and Nicki Tiffin [2,11,12]*

[1]Division of Human Genetics, University of Cape Town, Cape Town, South Africa. [2]Department of Integrative Biomedical Sciences, Computational Biology Division, University of Cape Town, Cape Town, South Africa. [3]Ministry of Health, Conakry, Guinea. [4]Chair of Public Health, University of Conakry, Conakry, Guinea. [5]Department of Pathology, Division of Chemical Pathology, Faculty of Medicine and Health Sciences, Stellenbosch University, Stellenbosch, South Africa. [6]Tygerberg and National Health Laboratory Service, Tygerberg Hospital, Parow, South Africa. [7]Wellcome Genome Campus Advanced Courses and Scientific Conferences, Hinxton, UK. [8]Centre for Medical Ethics & Law, Department of Medicine, Faculty of Health Sciences, Stellenbosch University, Stellenbosch, South Africa. [9]University of Yaounde 1 and Cameroon Bioethics Initiative (CAMBIN), Yaounde, Cameroon. [10]Immunology Department, School of Medicine, College of Health Sciences, Moi University, Kesses, Kenya. [11]Wellcome Centre for Infectious Disease Research in Africa, Institute for Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa. [12]Centre for Infectious Disease Epidemiology and Research, Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa. [13]These authors contributed equally: Victoria Nembaware, Katherine Johnston. *e-mail: nicki.tiffin@uct.ac.za

References

1. Popejoy, A. B. & Fullerton, S. M. *Nature* **538**, 161–164 (2016).
2. Nordling, L. *Nat. News* **544**, 20 (2017).
3. Adebamowo, S. N. et al. *Glob. Health Action* **11**, 1419033 (2018).
4. Tiffin, N. *Per. Med.* **11**, 187–196 (2014).
5. Wright, G. E. B., Koornhof, P. G. J., Adeyemo, A. A. & Tiffin, N. *BMC Med. Ethics* **14**, 21 (2013).
6. Paré Toe, L. et al. *PLoS One* **8**, e80800 (2013).
7. Hofmann, B. J. *Med. Ethics* **35**, 125–129 (2009).
8. Moodley, K., Sibanda, N., February, K. & Rossouw, T. *BMC Med. Ethics* **15**, 4 (2014).
9. Schroeder, D., Cook, J., Hirsch, F., Fenet, S. & Muthuswamy, V. *Ethics Dumping: Case Studies from North-South Research Collaborations* (Springer, 2017).
10. Tiffin, N. *BMJ Glob. Health* **3**, e001249 (2018).
11. Campbell, M. M. et al. *PLoS One* **12**, e0188466 (2017).
12. Munung, N. S. et al. *J. Med. Ethics* **42**, 132–137 (2016).
13. Budin-Ljøsne, I. et al. *BMC Med. Ethics* **18**, 4 (2017).
14. Kaye, J. et al. *Eur. J. Hum. Genet.* **23**, 141–146 (2015).
15. Igbe, M. A. & Adebamowo, C. A. *BMC Med. Ethics* **13**, 27 (2012).
16. Moodley, K. & Singh, S. *BMC Med. Ethics* **17**, 57 (2016).
17. Kim, H. et al. *J. Am. Med. Inform. Assoc.* **24**, 380–387 (2017).
18. Wright, G. E. B., Adeyemo, A. A. & Tiffin, N. *Hum. Genomics* **8**, 18 (2014).
19. H3Africa consortium array available soon. *Illumina.* https://www.illumina.com/company/news-center/feature-articles/h3africa-consortium-array-available-soon-.html (2016).
20. Browne, J. L. et al. *Trop. Med. Int. Health* **24**, 264–279 (2018).
21. Ogunrin, O., Woolfall, K., Gabbay, M. & Frith, L. *PLoS One* **13**, e0195171 (2018).
22. Rothstein, M. A. *Am. J. Bioeth.* **10**, 3–11 (2010).
23. Erlich, Y., Shor, T., Pe'er, I. & Carmi, S. *Science* **362**, 690–694 (2018).
24. Gymrek, M., McGuire, A. L., Golan, D., Halperin, E. & Erlich, Y. *Science* **339**, 321–324 (2013).
25. National Institutes of Health. *Update to NIH Management of Genomic Summary Results Access* Notice no. NOT-OD-19-023 https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-023.html (2019).
26. National Heart, Lung, and Blood Institute working group. et al. *Circ. Cardiovasc. Genet.* **3**, 574–580 (2010).
27. Dyke, S. O. M. et al. *PLoS Genet.* **12**, e1005772 (2016).
28. Lin, Y. et al. Development of a BFO-based informed consent ontology. in *Proc. 5th International Conference on Biomedical Ontology* (ICBO, 2014).
29. Genomic data toolkit. *Global Alliance for Genomics and Health* https://www.ga4gh.org/genomic-data-toolkit/ (2019).
30. Woolley, J. P. et al. *NPJ Genom. Med.* **3**, 17 (2018).
31. Harris, P. A. et al. *J. Biomed. Inform.* **42**, 377–381 (2009).
32. Folayan, M. O., Oyedeji, K. S. & Fatusi, O. A. *Dev. World Bioeth.* **15**, 1–7 (2015).
33. Kengne-Ouafo, J. A. et al. *BMC Public Health* **14**, 1026 (2014).
34. Moodley, K. & Beyer, C. *Biopreserv. Biobank.* https://doi.org/10.1089/bio.2018.0136 (2019).
35. Freudenthal, E. & Hecketsweiler, C. Où sont les échantillons sanguins infectés par Ebola? *Le Monde* https://www.lemonde.fr/afrique/article/2019/01/22/ebola-l-utilisation-opaque-des-echantillons-sanguins_5412571_3212.html (22 January 2019).
36. Freudenthal, E. Ebola's lost blood: row over samples flown out of Africa as 'big pharma' set to cash in. *The Telegraph* https://www.telegraph.co.uk/global-health/science-and-disease/ebolas-lost-blood-row-samples-flown-africa-big-pharma-set-cash/ (6 February 2019).
37. Community mistrust worsening DR Congo Ebola outbreak: study. *Al Jazeera* https://www.aljazeera.com/news/2019/03/community-mistrust-worsening-dr-congo-ebola-outbreak-study-190328062938247.html (28 March 2019).

Author contributions

N.T. conceptualized the framework and wrote the manuscript outline. N.T. and all authors developed the content of the framework and completed and reviewed the manuscript content.

Competing interests

M.K. is a non-executive director and shareholder of Gknowmix (Pty) Ltd.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0520-x.

nature research

Corresponding author(s): Prof. Nicki Tiffin

Last updated by author(s): Sep 5, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☐	A description of all covariates tested
☐	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☐	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☐	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☐	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No data were collected
Data analysis	No data analysis was performed

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

N/A

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	N/A
Data exclusions	N/A
Replication	N/A
Randomization	N/A
Blinding	N/A

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☐ Antibodies
☐	☐ Eukaryotic cell lines
☐	☐ Palaeontology
☐	☐ Animals and other organisms
☐	☐ Human research participants
☐	☐ Clinical data

Methods

n/a	Involved in the study
☐	☐ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Antibodies

Antibodies used	N/A
Validation	N/A

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	N/A
Authentication	N/A
Mycoplasma contamination	N/A
Commonly misidentified lines (See ICLAC register)	N/A

Palaeontology

Specimen provenance	N/A
Specimen deposition	N/A
Dating methods	N/A

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	N/A
Wild animals	N/A
Field-collected samples	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about [clinical studies](#)
All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	N/A
Study protocol	N/A
Data collection	N/A
Outcomes	N/A

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	N/A
Files in database submission	N/A
Genome browser session (e.g. [UCSC](#))	N/A

Methodology

Replicates	N/A
Sequencing depth	N/A
Antibodies	N/A
Peak calling parameters	N/A
Data quality	N/A
Software	N/A

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	N/A
Instrument	N/A
Software	N/A
Cell population abundance	N/A
Gating strategy	N/A

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	N/A
Design specifications	N/A
Behavioral performance measures	N/A

Acquisition

Imaging type(s)	N/A
Field strength	N/A
Sequence & imaging parameters	N/A
Area of acquisition	N/A

Diffusion MRI   ☐ Used   ☐ Not used

Preprocessing

Preprocessing software	N/A
Normalization	N/A
Normalization template	N/A
Noise and artifact removal	N/A
Volume censoring	N/A

Statistical modeling & inference

Model type and settings	N/A
Effect(s) tested	N/A

Specify type of analysis: ☐ Whole brain   ☐ ROI-based   ☐ Both

Statistic type for inference (See Eklund et al. 2016)	N/A
Correction	N/A

Models & analysis

n/a	Involved in the study
☐ ☐	Functional and/or effective connectivity
☐ ☐	Graph analysis
☐ ☐	Multivariate modeling or predictive analysis

Functional and/or effective connectivity	N/A
Graph analysis	N/A
Multivariate modeling and predictive analysis	N/A