07 December 2013~~04 November 2013~~

H3AFRICA CONSORTIUM DATA SHARING, ACCESS AND RELEASE POLICY

Table of Contents

Background

The goal of the Human Heredity and Health in Africa (H3Africa) Initiative(http://h3africa.org/) is to enhance the capacity of African researchers. H3Africa aims to undertake cutting edge research to advance understanding of the genetic and environmental determinants of common diseases in Africa and to use this knowledge to improve the health of African populations. The H3Africadata sharing, access and release policy is built upon H3Africa principles of ethics, governance and resource sharing that have been established by the NIH and Wellcome Trust (http://h3africa.org/ethics_governance_resourcesharing.cfm) and the H3Africa consortium members. These principles aim to strike an appropriate balance in ensuring that adequate safeguards are in place to protect participants, while maximizing the ability of investigators to advance research.

Principles for H3Africa Initiativedata sharing, access and release policy include:

- Maximizing the availability of research data, in a timely and responsible manner.
- Protecting the rights and privacy of human subjects who participated in research studies.
- Recognizing the scientific contribution of researcherswho generated the data.
- Considering the nature and ethics of the research proposed in establishing the timely release of data, and mechanisms of data sharing.
- Promoting deposition of genomic data in existing community data repositories whenever possible

General Guidelines

The H3Africa Initiative is committed to providing research data generated by the H3Africa research projects to the entire research community. The data to be released includes genomic and phenotypic data from properly consented individuals. H3Africa will initially prioritise collaboration of genotypic and phenotypic data within the consortium before releasing the data to wider scientific community. In compliance with current international standards to protect participant's identification, the H3Africa-generateddatawill be available to qualified researchers within the wider scientific community through a controlled access process at the European Genotype-phenotype Archive (EGA). Note, sequence data may also be deposited at the European Nucleotide Archive (ENA).To gain access, outside users must submit a request to use the data, which is then subject to approval byan independent H3Africa Data Access Committee (IDAC). After a data access request is approved, the outside user must sign a Data Access Agreement (DAA), which governs the terms and conditions on which access will be granted.

Specific Guidelines

Data Sharing and Release

Data types to be generated by H3Africa consortium:

- Geneticvariation data –(whole-genome or exome sequence data or genotyping data from arrays, as well as data from next-generation technologies and approaches)
    - All primary sequence data
    - Final variant calls passing Quality Control (QC) from sequence data
    - Genotypes from arrays, after QC
- Phenotype data - Phenotype information associated with the genotype data above.

The process and timelines are shown in figure 1.

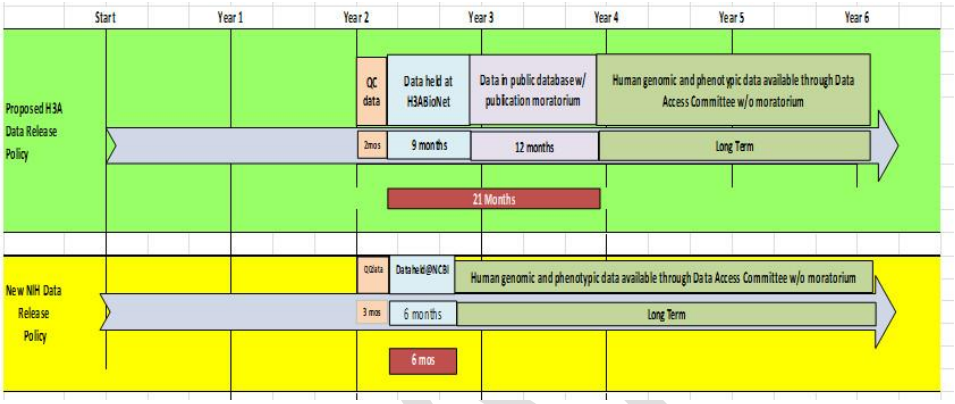

Figure 1. Proposed timelines for data submission and access for H3Africa, and comparison with the new NIH data release policy.

Whole genome or exome sequence data and DNA sequence variants that have been quality controlled and validated, will be sent to and deposited in H3ABioNet within approximately 2 months of generation. Although 2 months is the recommended time limit, the actual timelines and dates of deposition of the QC'ed data will be discussed ~~determined~~ between the project PI and their grant officer and plan will be reviewed by the appropriate funding agency.~~.~~ They will also determine what constitutes a complete dataset for submission.See Data Quality Control and Validation Guidelines (Appendix A) and H3ABioNet Data Submission Guidelines(Appendix B) for more information on the QC process and the process for submission of data to H3ABioNet.

Comment [GM([1]: Please review this and let me know what you think-this is only a recommendation and suggestion. To make sure there is some consistency among projects I added that the plans would be reviewed by funding agency as Jane or Audrey. Let me know what you all think.

Phenotypic data (clinical data or any other data associated with the clinical sample) will be sent to and deposited into H3ABioNetat the same time that the sequence data or array is submitted to H3ABioNet.See H3ABioNet Data Submission Guidelines (Appendix B).

Quality controlled genomic data and phenotypic data will be held in H3ABioNet for nine months.During this nine month period:

- H3ABioNet will format the data in preparation for submission to EGA. See EGA Data Submission Guidelines(Appendix C).
- H3Africa data generatorscan analyse and publish their data sets and analysis.
- H3Africa data generators should submit updates of their data to H3ABioNet

- H3Africa consortium members can negotiate for access to the data with the H3Africa consortium members who generated the data. Collaborative plans should be discussed and potential publications should be discussed with the H3Africa Publications Committee.

Following the nine month data hold at H3ABioNet, data will be sent to and publicly released at the EGA.

- Data sets deposited into EGA will become publicly available to bona fide scientists who request access through an independent H3Africa DACwith a twelve month publication moratoriumon the data set(see Data Access Committee Guidelines and Data Access Agreement).

PublicationMoratorium

To protect the publication rights of the investigators who generate data in H3Africa, a publication moratorium will be used. The moratorium will grant investigators exclusive right to publish analyses of the dataset for one year following the release of genotype or sequence dataset through the EGA, including the raw, pre-analyzed data. During this period of exclusivity, access to the data will be granted to other investigators through the DAC, who may analyze the data, but are expected not to submit their analyses or conclusions for publication during the exclusivity period. Publication exclusivity is expected to extend to all forms of public disclosure, including meeting abstracts, oral presentations, and publicly accessible electronic submissions (e.g. websites, web blogs).

Data will no longer be subject to the publication moratorium if the data has been published, or if one year has passed since the dataset used for analysis was made available through the EGA.

Details of the publication moratorium will be made available along with the datasets deposited in the EGA. Potential data users will be encouraged to contact the H3Africa project if there is any doubt about how the data should be used. Email templates will be provided to encourage this dialogue between potential data users and the H3Africa project. This dialogue will also demonstrate to journal editors whether data has been used appropriately.

It is expected that investigators who access H3Africa data acknowledge the project appropriately in any oral or written presentations, disclosures, or publications of the analyses.

Comment [GM([2]: This is for access during the 12 month moratorium and post moratorium access, I thought so.

I know we have talked about this but will the 12 month moratorium be managed by EGA with estimated start and end dates.

European Genome-Phenome Archive (EGA)

The EGA is a service for permanent archiving and sharing of all types of personally identifiable genetic and phenotypic data resulting from biomedical research projects (https://www.ebi.ac.uk/ega/). Information on EGA submission guidelines is provided in Appendix C. The EGA contains exclusive data collected from individuals whose consent agreements authorize data release only for specific research use or to bona fide researchers. Strict protocols govern how information is managed, stored and distributed by

the EGA project. Once processed, all data are encrypted for dissemination and the encryption keys are delivered offline.

The EGA will create new or update existing accounts only from the direction of the H3Africa DAC. The EGA will also require authorization from the corresponding Consortium Data Access Committee (CDAC) when the original application must be updated.

Approved data users are provided with an EGA account that includes personal details of the applicant, such as the name, postal address, email and it lists all accession rights granted by the different consortia and DAC. The account exists solely within the EGA system. The EGA account is for personal use only; the terms and conditions of the account prohibit sharing of the account details. The users will be able to request encryption keys, access authorized data files or request FTP/Aspera accounts from the system by using the tools integrated to the account.

Data Access

Internal Access

Access to the H3Africa data will distinguish between internal and external users. Internal users are members of the H3Africa consortium. Consortium members include project PIs, their co-investigators and student and staff members. Any other person, including collaborators of PIs (not involved directly in H3Africa) would be considered external users. Internal access should be requested directly from the PI who generated the data. Once access is granted, the PI can choose to send the data directly to the requestor, or can ask H3ABioNet to provide access to the named internal user.
Note, internal access can be requested and granted during the 9 months that the data is held at H3ABioNet, or at any time after submission to the EGA.

Comment [GM([3]: Will the PI determine who is considered consortium members for this access?

Comment [GM([4]: Just to be clear for me, for consortium members to access data once it is at EGA, they do not have to apply and be reviewed by data access committee?

External Access

The basic descriptive and aggregate summary information generated by each H3Africa study will be available publicly through the EGA. Access to the genotype and phenotype datasets submitted and stored in EGA, along with appropriate automated calculations (e.g., quality control measures, simple genotype-phenotype associations, or a listing of all variants known to be in linkage disequilibrium with variants measured in the genotype), will be provided for research purposes through an independent DAC.

Potential data users will be required to provide a summary of their proposed research.See Data Access Request Form (Appendix E).The review of requests for data access will include screening for inappropriate use of the data, to ensure that it complies with the ethical consent associated with the data or samples. This will be based on proposed area of research, not on scientific strength or novelty. There will be no scientific-peer review of the proposed research and duplicative analyses will be acceptable.

Approved data users will be required to agree to the Terms and Conditions of a Data Access Agreement (DAA), which aims to protect the privacy and interests of the research participants. The DAA will require users to agree to:

a) Use the data only for the approved research;

b) Protect data confidentiality;
c) Follow appropriate data security protections;
d) Follow all applicable laws, regulations and local institutional policies and procedures for handling genotype, sequence and phenotype data;
e) Not attempt to identify individual participants from whom data within a dataset were obtained;
f) Not sell any of the data elements from datasets obtained from the EGA;
g) Not share with individuals other than those listed in the request any of the data elements from datasets obtained from the EGA;
h) Agree to the listing of a summary of approved research uses within the EGA along with his or her name and organizational affiliation;
i) Agree to report, in real time, violations of the H3Africa Data Access policy to the DAC;
j) Acknowledge the H3Africa policy with regard to publication and intellectual property;
k) Respect first publication rights as described in the embargo;
l) Cite the original H3Africa publication that describes the analysis of the dataset(s) used, and;
m) Provide annual progress reports on research using the genotype, sequence and or/phenotype dataset.

A list of projects for which access has been approved (PI and title) will be made publicly available on the H3Africa website. Records will be kept of who gained access, and when, to ensure that the terms of access and the conditions of the publication moratorium are complied with.These records will not be made publicly available, but will be available to the H3Africa Steering Committee for oversight purposes.

Data Access Committee

Data access requests will be managed through an independent Data and Sample Access Committee (DAC) by appointed members with relevant expertise in areas such as the relevant particular scientific disciplines, research participant protection, and privacy (Appendix D). Investigators and institutions seeking data from the EGA will be expected to meet data security measures (such as physical security, information technology security, and user training) and will be asked to submit a data access request, that is co-signed by the investigator and the designated Institutional Official(s). Data access requests should include a brief description of the proposed research use of the requested dataset(s).

The DAC will be responsible for receiving and reviewing applications for access to H3Africa data and determining if:

- The principal investigator is a bona fide researcher;
- The research use is acceptable;
- The DAA has been completed satisfactorily and includes all required signatures.

The DAC or their designees will review requests for access to determine whether the proposed use of the dataset is scientifically and ethically appropriate and does not conflict with constraints or informed consent limitations identified by the institutions that submitted the dataset to the EGA. In the event that requests raise concerns related

to privacy and confidentiality, risks to populations or groups, or other concerns, the DAC will consult with other experts as appropriate.

***Release of patient/donor identifying data**:*

H3Africa may release curated and coded phenotype, exposure, genotype, and pedigree data. To protect the rights and privacy of human subjects who participate in the studies, clinical metadata, genomic, or other data sets, or a subset of the clinical and other metadata that may potentially identify human subjects of samples **shall not** be released in openly accessible public databases. These detailed data will be made available through a controlled access process through the EGA.

Clinical metadata and other fields that may potentially uniquely identify an individual will be carefully reviewed and flagged prior to sharing and releasing any clinical metadata to openly accessible public databases.

In order to minimize risks to study participants and their communities, data submitted to the EGA will be de-identified and coded using a random, unique code to ensure the identities of data subjects cannot be readily ascertained or otherwise associated with the data by the repository staff or secondary data users.Further details are provided in the EGA Submission Guidelines (Appendix C) and Data Access Agreement (Appendix F).

Monitoring

A breach of any of the conditions of the Data Access Agreement (DAA), including failure to respect the publication moratorium, will terminate the DAA and access to the data will be withdrawn. Future access mayalso be denied to individuals found responsible for a previous breach of the conditions of the DAA. In addition, when a breach of the publication moratorium is suspected or takes place, the H3Africa Steering Committee and/or the DAC may contact the appropriate journal editor with evidence that data use conditions have been breached and to request that any manuscripts be withdrawn.

Data users will be expected to submit annual reports to the DAC describing the use of the data to ensure that it complies with the terms of the DAA and the provisions laid out in the associated informed consent.

Intellectual Property

It is the aim of H3Africa that genotype-phenotype data made available through the EGA and all conclusions derived directly from them will remain freely available, without any licensing requirements, for uses such as, but not necessarily limited to, markers for developing assays and guides for identifying new potential targets for drugs, therapeutics, and diagnostics.H3Africa discourages any premature claims on pre-competitive information that may impede research, though it encourages patenting of technology suitable for subsequent private investment that may lead to the development of products that address public healthcare needs.

The filing of patent applications and/or the enforcement of resultant patents in a manner that might restrict use of H3Africa genotype-phenotype data could diminish the potential public benefit they could

provide. Approved users and their institutions, through the execution of a Data Use Certification, will acknowledge the goal of ensuring the greatest possible public benefit from H3Africa data.

7. Intellectual Property

By requesting access to genomic dataset(s), the Requester and Approved Users acknowledge the intent of the NIH that anyone authorized for research access through the attached Data Access Request follow the intellectual property principles within the NIH GWAS Policy for Data Sharing as summarized below:

> Achieving maximum public benefit is the ultimate goal of data distribution through the NIH genomic data repositories. The NIH believes that these data should be considered as pre-competitive, and urges Approved Users to avoid making IP claims derived directly from the genomic dataset(s). However, the NIH also recognizes the importance of the subsequent development of IP on downstream discoveries, especially in therapeutics, which will be necessary to support full investment in products to benefit the public.
>
> It is expected that these NIH-provided data, and conclusions derived therefrom, will remain freely available, without requirement for licensing. The NIH encourages broad use of genomic datasets coupled with a responsible approach to management of intellectual property derived from downstream discoveries in a manner consistent with the NIH's Best Practices for the Licensing of Genomic Inventions and the NIH Research Tools Policy .

Comment [GM([5]: This is text from the NIH GWAS documents and I will need to show H3Africa text and NIH text to our NIAID Tech Transfer folks. I think it says the same thing but will need it reviewed.

Appendices

A. Data Quality Control and Validation Guidelines

B. H3ABioNet Data Submission Guidelines

C. EGA Submission Guidelines

D. Data Access Committee Guidelines

E. Data Access Request Form

F. Data Access Agreement

G. Policy Statement& Institutional Certification