Submission of WHI Genetic Data to dbGaP by Ancillary Study Investigators

QUICK INFO SHEET

(Also see more detailed "Submission of WHI Data to dbGaP – Guidelines")

Summary: If an Ancillary Study is submitting its data as a WHI Substudy, the procedure is to obtain Institutional certification, register the substudy with dbGaP, get a link file with Common-dbGaP Subject ID-dbGaP Sample IDs from the WHI-CCC, create a Subject-Sample mapping, and submit data to dbGaP via the submission portal. Data sharing and the informed consent status of participants whose data may be deposited in dbGaP is done in consultation with WHI, since consent for genetic studies does not directly imply consent to submit genotype and phenotype data to a more public database such as dbGap.

1) WHI data should be submitted to dbGaP as a substudy to WHI (phs000200).

a) PI obtains Institutional Certification (IC), as depositor of the data:
   i) Use template IC: https://gds.nih.gov/Institutional_Certifications.html
      (1) Template for WHI data is "For Studies Using Data Generated from Cell Lines Created or Clinical Specimens Collected BEFORE January 25, 2015" – – "that have consent"
      (2) WHI consent group titles and modifiers are:
         (a) Health/Medical/Biomedical-IRB (HMB-IRB)
         (b) Health/Medical/Biomedical-IRB-Non-profit-use (HMB-IRB-NPU)
      (3) Data Use Limitations (DULs) must be listed*
         (a) Print the IC, cut and paste DULs in before the page 3 signatures and rescan and sign to have a complete form, or
         (b) Attach as an addendum, as long as it is clear that the signature covers the entire package.
      (4) Genomic Summary Results (GSR) is 'controlled access' (sensitive designation), with explanation: Genomic Summary Results should be maintained under controlled access as some of the data includes relatively small minority populations that are vulnerable and have been stigmatized in the past, as well as some sensitive phenotypes."

   ii) NOTE: Ancillary Study IRB cannot override the WHI parent consent group categories/DUL.

b) PI works with the NHLBI Genomic Program Administrator (GPA) to register the substudy:
   Alex Runko/Sue Chen/DAC staff nhlbigeneticdata@nhlbi.nih.gov
   i) Submit study registration form specific to the substudy
   ii) Submit Institutional certificate (IC)
   iii) Submit Acknowledgements to WHI-CCC (WHI-CCC keeps master list and updates dbGaP).
   iv) Universal WHI Data Use Certificate (DUC) will automatically be used. No effort is needed by Ancillary Study PIs.

c) PI submits files via the submission portal after registration is complete
   i) An automatic submission invitation and instructions are sent to the substudy PI after the study is registered, and, if specified, secondary contact/data submitter.
   ii) Submit study config document and other required files via the dbGaP submission portal.
      (1) Refer to main WHI top level study in study config as much as possible to avoid duplication and make updates easier to manage (e.g., WHI study description, WHI study timeline).
   iii) The lab data you submit to dbGaP should be linked **only** to the WHI dbGaP Subject ID (ID used to uniquely identify a WHI participant's data [phenotype or genotype] on dbGaP).
      (1) WHI blind duplicates do not have associated covariate data and are not to be submitted to dbGaP.
      (2) ***The link between the dbGaP Subject ID and the WHI Common ID or Draw/Vial ID is never to be submitted to dbGaP. The WHI Common ID or Draw/Vial ID is never to be submitted to dbGaP.***
   iv) Notify the WHI-CCC when registration is complete and the substudy accession number is assigned.

2) Requests for an exception to be a 'stand-alone' study per a condition of the funding agency (e.g., consortium studies with several cohorts).

a) PI confirms with PD/PO of funding award and submits request to the WHI Project Office.
b) PI ensures correct consent groups and DULs are applied.
c) PI notifies WHI-CCC of phs number once registered.
d) Refer to WHI as much as possible
   i) E.g., WHI study description, WHI study timeline in Study config.
   ii) Standalone studies should ensure the correct acknowledgements for WHI are used.

I. Glossary of Terms (work in progress)

GPA = Genomic Program Administrator. Person responsible at dbGaP for registering the study.
dbGaP Curator (phenotype and genotype) = person(s) responsible for QC of submitted phenotype and genotype data.
PD/PO = Project Director/Project Officer of the funding award. Provides oversight to dbGaP submission.
DUL = Data Use Limitation. Listed on template Institutional Certification (IC) per IRB review of the data to be posted by the submitting institution.
DUC = Data Use Certification.
DAC = Data Access Committee. Each institute has their own DAC and DAC chair, responsible for reviewing data access requests and ensuring the use of the data complies with the data use limitations.
DAR = Data Access Request. Application to request data from dbGaP.
GSR = Genomic Summary Results. summary statistics that may be computed by an NIH-designated data repository, to include systematically computed statistics such as, but not limited to: 1) allele frequency information (e.g., genotype counts and frequencies, or allele counts and frequencies); and 2) association information (e.g., effect size estimates and standard errors, and p-values). **In May 2019, WHI Leadership determined that GSR should be designated as sensitive and kept under controlled access.**

II. Attachments/Appendix:

WHI template IC
WHI data use limitations as they appear on dbGaP
WHI DUC
WHI acknowledgements
WHI study config – example
WHI IDs

*WHI Data Use Limitations;

"All requestors and data users must provide a local Institutional Review Board (IRB) letter from their respective institution approving conduct of the proposed research project. Only full or expedited approvals will be accepted. There are two consent groups for this study:
a) Health/Medical/Biomedical – IRB, and b) Health/Medical/Biomedical – IRB for use by not-for-profit organizations only.

Use of the WHI data to conduct non-genetic research is prohibited. WHI data may not be used to investigate individual pedigree structures, individual participant genotypes, perceptions of racial/ethnic identity, variables that could be considered as stigmatizing an individual or group, or issues such as non-maternity. Users acknowledge that American Indian participants were self-identified and were primarily dwelling in urban areas and are not representative of the diverse American Indian population across the United States. Users agree not to use the data to infer tribal status or affiliation."