WHI Policy for Direct Access to WHI Genetic Data

Background

WHI participants have been included in multiple genome-wide, sequencing, exome-wide, and other genotyping efforts. The FHCRC IRB determined in 2009 that the release of genetic data for these participants through public websites is allowable, and in 2011 that the release of linked phenotype data is allowable. The distribution of WHI genetic data is handled through a 3rd party (dbGaP). To promote collaborative research with existing WHI principal investigators, and at the request of the NHLBI project office, we have added a direct means for WHI Sponsoring Investigators (WHI PIs) to access WHI genetic data per the procedures outlined below. WHI PIs may designate a downloader who can receive and work with the data. The downloader does not need to be at the same institution as the WHI PI. In some cases, other investigators interested in using this resource may be eligible with a WHI PI to sponsor their data use (e.g., when the data is not yet available on dbGaP, or when linking data to existing Ancillary Study results). The WHI PI may need to provide an email acknowledging they are aware of the genetic data transfer.

For information about available WHI genetic datasets that may be made available through the procedures below, please see "GWAS Data and Blood Specimen Results" on the WHI SharePoint site.   Phenotype data will not be provided through this mechanism, but can be downloaded from the WHI sharepoint site (as appropriate per the signed data use agreement).

Procedure for data access

1. Seek P&P or ASC approval for the use of WHI genetic data . The investigator should specify which components of the genetic data will be included in the analysis. This can be listed as SNPs, genes, pathways, chromosomal regions, chromosomes, whole genome, etc.

2. Seek IRB approval for the use of genetic data (full or expedited only). This IRB approval may occur before or after the study has been reviewed and approved by the WHI committee, but should occur prior to downloading the genetic data.

3. Submit a copy of the WHI Data Use Agreement (DUA) signed by the investigator and all analysts working directly with the data.

4. Contact the WHI helpdesk (helpdesk@whi.org). Your request will be forwarded to WHI Statistical staff to prepare the appropriate data for download.

5. WHI Statistical staff will transfer the appropriate file(s) to Aspera (free software, also used for this purpose by dbGaP), and use Aspera to send the Investigator notification that the files are ready for download.

    i. Space needed for the whole set of compressed imputed GWAS data is 2 terabytes. Data is broken into files by study and chromosome, but generally not further (smaller subsets of SNPs will not be extracted for Investigators). Data is in NetCDF-4 format and WHI staff will provide a help document for use of these files using R to Investigators. Data must be downloaded within 7 days, and can only be downloaded 1 time using the link (for security purposes) in the email sent by WHI staff.

    ii. Because of the size, we can only distribute to one investigator at a time. Download takes approximately 36 hours with high-speed internet connections. Due to these practical limitations, there may be a queue for data access, and it is possible that the waiting period to receive data may be several weeks.

Annually, investigators who have received genetic data to date will receive an email requesting an informal status update of the file (in use, no longer in use, destroyed), along with a reminder to not further disseminate the data and to destroy the data if it is no longer in use.

The WHI-CCC will submit to IRB annually a tracking sheet listing each request, to include PI name, institution, MS# or AS#, confirmation of receipt of investigator's local IRB and data use agreement (if applicable), and the annual status update.

Note that as with all WHI manuscripts, the WHI Publications and Presentations Committee (P&P) will review manuscripts prior to journal submission. As with dbGaP data, publication of data must be censored or aggregated for cells with < 5 participants.