dbGaP Request Procedures to Access Individual-Level Data

Developed and operated by the National Library of Medicine's National Center for Biotechnology Information (NCBI), dbGaP archives and distributes data from studies that have investigated the relationship between phenotype and genotype, such as genome-wide association studies (GWAS). The database provides two levels of access: open (available to anyone with no restrictions), and controlled (requiring preauthorization). The controlled-access portion of the database provides for downloads of individual-level genotype and phenotype data that have been de-identified (i.e., no personal identifiers, such as name, etc.).

In order to request access to any of the individual-level datasets within the controlled-access portions of the database, the Principal Investigator (PI) and the Signing Official (SO) at the investigator's institution will need to co-sign a request for data access, which will be reviewed by an NIH Data Access Committee at the appropriate NIH Institute or Center. In order to complete this step, which utilizes the SF 424 (R&R) form, both the PI and SO will need to have accounts with the NIH eRA Commons. These are the same accounts used to apply for grants, and PIs and SOs who already have such accounts do not need to do anything further to make them applicable to the dbGaP controlled-access authorization process. Information on applying for an eRA Commons account can be found at https://commons.era.nih.gov/commons/

Assuming eRA accounts are in place, the process for requesting access can be started from two places. When browsing the dbGaP website of public content, users may review the open-access information on available projects at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap to determine which datasets are applicable to the user's research questions; the description pages for each study include links that will direct the user to the controlled-access data request login page. Alternatively, when the user already knows which datasets will be requested, he or she may start directly at the controlled-access login page http://dbgap.ncbi.nlm.nih.gov/aa/wga.cgi?page=login .

The PI should follow the "Login" link on the right side of the controlled-access login page; she will be presented with the standard "NIH LOGIN" page and asked to fill in her eRA name and password; if it is the PI's first dbGaP request, she will be taken to a "preferences" page to fill out contact information; the PI clicks on

the “my projects” tab, where a link is provided for new data requests; the PI then follows the provided directions for completing the 424 (R&R).

Among the information the PI will provide in these forms is the name of the preferred SO (registered SOs at the PI’s institution are pre-listed on the form), a statement summarizing the proposed research use for the requested data, and a list of collaborating investigators at the same institution.

Important! Collaborators at other institutions will need to submit separate requests for co-submission with their local SOs

Submission of the data access request will constitute agreement and acknowledgment by both the PI and the SO to the terms of use for the specific dataset(s) requested, which are detailed in the accompanying “Data Use Certification” (DUC) documents that are provided on the website.
The DUC statements outline policies and procedures for using the data, such as limiting use to the project described in the Data Access Request form; not distributing the data beyond those permitted to handle it; not attempting to identify or contact study participants from whom phenotype data and DNA were collected; awareness of the specified principles regarding intellectual property; adhering to policies on the timeframe for publications stemming from the data; and other provisions designed to protect the confidentiality of study participants and to foster scientific advance.

After the PI completes the electronic data request process, the SO will be notified by email that a request has been submitted and is awaiting his signoff. The SO then uses his eRA name and password to enter the dbGaP authorized access system, where he will be presented with the PI’s application to review. The SO has the options of editing the forms, returning the forms to the PI for revision, or signing off that the submitted application is valid. To help ensure applications move through the submission and review process in a timely way, the SO and PI will receive various emails updating them on the status of a request or any required actions. The data access request is then reviewed by the appropriate Data Access Committee(s) at NIH, and both the PI and SO will be notified by email of approval or disapproval.

More information on dbGaP and links to the information pertaining to each available dataset can be found at
http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap