Downloading dbGaP SRA Data**

Sequence Read Archive Submissions Staff

**Please Note: The information provided in this section regarding the download of individual-level data from dbGaP assumes that you have completed the dbGaP data access request process, and have been granted authorized access to individual level data.

If you do not have authorized access to dbGaP individual-level data, accessing individual-level data using the instructions in this section will not be possible, as you will not have access to the appropriate password-protected dbGaP sites mentioned. For information on how to apply for authorized access, please see the "Data Access Request" section of the dbGaP FAQ Archive.

I found some human SRA data that I want to download, but the entry for the data I want doesn't have direct links to download the data —I have to go to dbGaP authorized access. Why can't I just download the data directly?

The sequence data you found is human individual level data —this means that **the data was derived from the tissue of a particular person who signed an informed consent agreement which specified restrictions in the way the data derived from their tissue can be used**.

Because of these restrictions, **the data can only be accessed by investigators who have successfully applied for dbGaP authorized data access**. This application process ensures that the stated research purpose of the applicant follows the restrictions stipulated by the study participant and that the access applicant and his or her institution will abide by the study's data use certification and terms of use.

Once you are granted authorized access:

- You will not have access to all the data housed in the authorized access portion of dbGaP since data access policies are determined on a per-study basis.
- Access will be only to the data you've been approved for in one particular study.
- You can request additional data sets from within the study you have been approved for, but these requests must be approved before the data can be released to you.
- If you need data files from more than one study, you will need to apply for authorized access for each study that generated the data you require.

You will find a brief description of the steps required access and download SRA data contained within dbGaP in this SRA Knowledge Base.

How do I access and download the dbGaP SRA data?

Below is a brief overview of the steps to access and download SRA data contained within dbGaP:

1. Find the dbGaP study and consent group names in any of the SRA report types (SRP, SRR, SRS) where the data you want appears. You will need these names for your controlled access application.

2. Apply for authorized access to the dbGaP study and consent group. An overview of the process for requesting access to a dbGaP study can be found here.
3. Request data sets from the study for which you have been granted access.
4. Obtain the dbGaP repository key file (".ngc" file). Information on .ngc files can be obtained here.
5. Download the SRA toolkit and use the included Java utility to configure a dbGaP workspace.
6. Run the Java configuration tool and import the .ngc file.
7. **IMPORTANT:** Switch, or 'cd' into the configured workspace directory made in (5) you can then download and operate directly on encrypted SRA data ("on-the-fly" download and "as-needed" decryption) or move a previously-downloaded ".sra.ncbi_enc" encrypted SRA file into your configured workspace and operate directly on it.

Note: **Only a** person classified as **a Principle Investigator or PI** (NIH extramural or intramural researcher) **can apply for dbGaP authorized access**.

Non-PIs cannot independently apply for access to individual-level data, but they can be approved for local access to downloaded data files within the PI's lab if they are listed as collaborating investigators on a PI's application.

An explanation of why these steps are necessary is available elsewhere.