



## Decrypting and Extracting Data

Created: October 22, 2008; Updated: August 12, 2013.

### File Decryption

#### Are downloaded files encrypted? If so, do I need to decrypt them and how?

The following instructions are nearly identical in all supported platforms.

##### 1. Different treatment of SRA and non-SRA data

The data files distributed through the dbGaP are all encrypted by NCBI's data encryption algorithm. These files have a file suffix ".ncbi\_enc", indicating that they are NCBI encrypted files. Not all encrypted data however need to be decrypted.

The SRA (short-read-archive) data distributed through the dbGaP are encrypted **but there is no need to decrypt them**. The NCBI SRA toolkit can work directly on encrypted SRA data without decryption. Decrypted SRA data is in a binary format that is not human readable and can only be processed by the SRA toolkit anyway.

You need NCBI SRA toolkit to work on SRA data. The SRA toolkit is a collection of utilities that can dump, extract, and convert SRA data to different data formats. The vdb-decrypt utility included in the SRA toolkit can be used to decrypt any encrypted dbGaP data.

The dbGaP data other than SRA (non-SRA data) need to be decrypted before use. If you are only working on non-SRA data, you can download the NCBI Decryption Tool, which is a sub-set of the SRA Toolkit. It only includes utilities related to data decryption. If you already have SRA toolkit setup, you don't need to download NCBI decryption tool because the vdb-decrypt utility is included.

Both NCBI SRA Toolkit and NCBI Decryption Tool are available from [here](#).

##### 2. The dbGaP repository key

dbGaP repository key is a dbGaP project wide security token required for configuring NCBI SRA toolkit and decryption tools. The key is provided in a file with suffix ".ngc". It can be obtained from two places in PI's dbGaP account.

1. The first place is the project page under "My Projects" tab, through a link named "get dbGaP repository key" in the "Actions" column. The key downloaded from here is valid to all downloaded data under the project.
2. The second place is the download page under "Downloads" tab, through a link named "get dbGaP repository key" in the "Actions" column.

##### 3. Toolkit Configuration and import repository key

The NCBI decryption tool is a subset of the SRA Toolkit. The steps of setting up both tools are nearly identical. In either case, a dbGaP repository key for the respective dbGaP project should be

downloaded from PI's dbGaP account, and the tool should be first configured using "vdb-config", a command line utility available under the "bin" directory of the toolkit. See [here](#) for detailed instruction.

#### 4. Decrypting Non-SRA Data

The Non-SRA data distributed through the dbGaP need to be decrypted before used for anything. The tool named "vdb-decrypt" under NCBI sra-toolkit or NCBI decryption Tools is for data decryption.

To decrypt non-SRA data, go to the dbGaP project directory (workspace) setup through the toolkit configuration, issue the following command from a command line: It is important to remember that the command line has to be run directly from the dbGaP project directory.

A typical vdb-decrypt command should be like this:

```
$ /path-to-your-sratoolkit-installation-dir/bin/vdb-decrypt --ngc /path-to-ngc-file-dir/xxxxx.ngc /path-to-top-level-download-dir/
```

#### 5. More about NCBI SRA Toolkit

Please refer to the [documentation of sra-toolkit](#) for more about various utilities available under the sra-toolkit.

(12/09/2020)

## SRA to BAM format conversion

**We would like to get the data in BAM format but they are only available in SRA format. What can we do?**

Most of the sequencing data available through the dbGaP are in SRA format. The SRA data can be converted to BAM format using the sam-dump combined with samtools. The sam-dump utility is available under the SRA toolkit. More information about the sam-dump is available at [here](#), and the information about the samtools can be found from [here](#).

(12/24/2013)

## SRA fastq-dump Utility

**How to convert downloaded SRA data into FASTQ format?**

Please visit the section related to the fastq-dump utility in [SRA Download Guide](#). If you have further questions regarding SRA (Short-Read-Archive) data, please directly contact NCBI's SRA group ([sra@ncbi.nlm.nih.gov](mailto:sra@ncbi.nlm.nih.gov)). They are better able to help with SRA related issues.

(10/19/2011)