

sratoolkit/2.10.7 to download dbGAP data

Due to Aspera license limitations, users have to install the software under their home to use the software.

- Start interactive job, and create working folder and load sratoolkit module
- Configure sratoolkit. Only need to do this once:
- Downloading dbGaP repository Key and upload to O2:
- Use sratoolkit prefetch, which try ascp then http, to download sra data, then convert the data from .sra to .fastq format

Start interactive job, and create working folder and load sratoolkit module

For example, for user abc123, the working directory will be

```
srunc --pty -p interactive -t 0-12:0:0 --mem 2000MB -n 1 /bin/bash

mkdir /n/scratch3/users/${USER:0:1}/${USER}/testDbGaP

cd /n/scratch3/users/${USER:0:1}/${USER}/testDbGaP

module load sratoolkit/2.10.7
```

Configure sratoolkit. Only need to do this once:

```
# Configure sratoolkit
vdb-config --interactive

# Directly press x key to quit

# By default, sratoolkit uses working directory as cache. It is better to
use scratch3 instead:
echo /repository/user/ad/public/root =
"/n/scratch3/users/${USER:0:1}/${USER}/ncbi\" >>
~/ncbi/user-settings.mkfg
```

Downloading dbGaP repository Key and upload to O2:

My Research Projects

You are listed as a designated downloader. You can see which datasets are available for download by clicking on the "My Requests" tab above.

#	Project	Actions
		run selector file selector launch browser net embargo report get dbGaP repository key

```
# Upload the dbGaP repository Key to O2:  
scp ~/Download/prj_phs710EA_test.ngc  
$USER@transfer.rc.hms.harvard.edu:~/ .ncbi
```

Use sratoolkit prefetch, which try ascp then http, to download sra data, then convert the data from .sra to .fastq format

```
# Load sratoolkit module  
module load sratoolkit/2.10.7  
  
# Use prefetch to download SRA file.  
prefetch --ngc ~/.ncbi/prj_phs710EA_test.ngc -p SRR1219902  
  
# Convert SRA file to FASTQ with fastq-dump. -X 5 means only gives first  
five reads to give a quick test.  
fastq-dump -X 5 --ngc ~/.ncbi/prj_phs710EA_test.ngc --split-files  
SRR1219902
```