

Downloading RNA-Seq datasets

Written by David Manahan, 2/2/18

Create your accession list.

- 1.) Identify the accession number assigned to the RNA-Seq dataset deposited in GEO. The number is usually found at the end of the paper, and generally starts with the letters 'GSE.' Ctrl + F function for 'deposit' should help your search.
- 2.) Go to GEO's website, <https://www.ncbi.nlm.nih.gov/geo/>. Enter the GEO accession number into the search bar. A results page should appear corresponding with your entered dataset.
- 3.) Scroll to the bottom of the webpage until you reach the 'Relations' section. Open the 'SRA' link (usually the link starts with the letters 'SRP').
- 4.) On this page, find the 'Send to:' drop down menu in the upper-right hand corner. Under 'Choose Destination' select 'File.' A new drop down menu will appear asking you to select a format. Choose 'Accession list' and then 'Create File.' This should download the list as 'SraAccList.txt'
- 5.) Open the 'hpc-transfer.usc.edu' node and navigate to your working directory, which should be located in the 'staging' directory. The home directory has very little room and likely will not be able to store all your files produced in a single analysis.

```
cd ~/path/to/working/directory
```

- 6.) In your working directory, create a sub-directory where you'll store all the raw data from online datasets and the outputs from the RNA-Seq analysis pipeline.

```
mkdir insert_directory_name
```

(*Feel free to replace "insert_directory_name" with a name of your choice*)

- 7.) Change your working directory to insert_name.

```
cd insert_directory_name
```

- 8.) Create a file that will be your accession list. In this file, you will list all the names of the sequencing data files found in your chosen dataset on GEO.

```
nano accession_list
```

- 9.) This will open a blank page. Now open 'SraAccList.txt' and copy its contents into accession_list.

- 10.) Save the file with 'ctrl+o' and confirm the filename by hitting return. Exit the file with 'ctrl+x'

- 11.) From your data directory, execute your downloads:

```
../downloads.sh
```

- 12.) Once the download is complete (i.e. when the the percent downloaded bar is done *and* you are able to enter commands again), exit out of hpc-transfer.usc.edu.

```
exit
```