

Short Read Workshop – Homework Day 3

1. In class today, we learned how to write an sbatch script and submit it with some basic parameters. Create a new directory in your home directory called “day_3_homework.” Create an sbatch script that will output the date, sleep for 10 minutes, and run date again to show the 10 minutes difference. Save the error and output files and script to show the instructors. a. HINT: sleep is a command you have not used yet. Look on the man page for it. date is also a command you may not have used.
2. Write a second sbatch script and download (using curl) some 1000 genomes data. If you don't want to pick some custom data, use `ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR792/SRR792473/SRR792473_1.fastq.gz`. If you're feeling adventurous, you can try to find your own data!
(http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/pilot_data/data/)

*NOTE: The downloads will very likely take longer than in class today since we're pulling much larger files than we did in class today. Be patient, be persistent, and you will be successful.

3. Run fastqc on the fastq.gz file. Make sure the error and output files look as you'd expect!
4. Rsync the html file from the fastqc command to your local machine, open it, and view the results.
5. Extra credit: use the date command at the before your curl and fastqc command for multiple different SRR files. Then look in standard out files and see how long it took you to download the SRR files today. Which one took more time? What are the file sizes of the SRR files? Do the time and the file size correlate?
6. Watch the videos for day 4