Sammy Mustafa

1. **wget from the ENA**
   a. Created awk file to extract the err numbers of interest from the ERP004763 sample mapping file, saved this in a txt file
   b. Used `grep` with the txt file of err numbers to filter the tsv with all the ftp addresses with only those of our samples of interest, saved this to our own file report tsv
   c. Extracted only the seventh column of this file report tsv to only keep the ftp address of the sequencing files we need to download, saved this as a shell script
   d. Ran this shell script
2. **fastq-dump from NCBI**
   a. cat my_errnums.txt | while read line; do fastq-dump --gzip $line; done
   b. ^^ uses each line of the txt file with the isolated errnums of interest to run into the fastq-dump and gzipping process
3. **fasterq-dump from NCBI, combined with gzip**
   a. cat my_errnums.txt | while read line; do fasterq-dump $line; gzip $line.fastq; done
   b. ^^ does same thing, also includes the process of gzipping the file
4. **ascp from the ENA**
   a. Used the modified awk with ascp script from the lectures with my sample's file report, saved this as a shell script and ran it
5. **globus (command line) from the ENA**
   a. awk '{print "/gridftp/ena/fastq/" substr($0,1,6) "/" $0 "/" $0 ".fastq.gz /projects/e31265/ssm9301/opt4/" $0 ".fastq.gz"}' my_errnums.txt
   b. ^^ script produces the "globus_batch_list.txt" of interest by printing the errnums of interest in both the source and target locations, does this for each errnum
   c. Initiate the transfer with the command including this txt file

	wget	fastq-dump	fasterq-dump	ascp	globus
Time (min)	206	73	60	13	1

Table of time to download the same sequencing data with different methods, transferred in the evening (6:20-9:46 PM)