Using Feature counts walk through

Step 1) Copy our scripts

Log in to the super computer
Make a directory /scratch/User/yourusername/day6/
Copy the two scripts we have provided in
`/scratch/Shares/public/sread2021/scripts/day6/`
Cd into the new directory

```
maryallen — maallen3@ip-172-31-38-192:/scratch/Users/maallen3/day6 — ssh maallen3@18.219.252.252 — 146×34
[maallen3@ip-172-31-38-192 ~]$ mkdir -p /scratch/Users/maallen3/day6/
[maallen3@ip-172-31-38-192 ~]$ scp /scratch/Shares/public/
genomes/   sread2021/
[maallen3@ip-172-31-38-192 ~]$ scp /scratch/Shares/public/sread2021/
algorithms/          cookingShow/         data_files/          homework_data_files/ README.txt           scripts/
[maallen3@ip-172-31-38-192 ~]$ scp /scratch/Shares/public/sread2021/scripts/day6/* /scratch/Users/maallen3/day6/
[maallen3@ip-172-31-38-192 ~]$ cd /scratch/Users/maallen3/day6/
[maallen3@ip-172-31-38-192 day6]$ ls
counts  featureCounts.R  sr_featcounts.sbatch  workshop
```

Step 2) Edit Our scripts

In the R script Change your working directory
In the sbatch script change your email and error and output files
####Make sure the error and output directory exist before you run!!!!!!

```
maryallen — maallen3@ip-172-31-38-192:/scratch/Users/maallen3/day6 — ssh maallen3@18.219.252.252 — 146×34
###### Running featureCounts
####### Author: Taylor Jones
####### Here we will learn how to download a package, what metadata table is (and why it is important),
####### and run featureCounts, which counts reads over genes.

#########################################################################################################################
####### We will want to start fresh and clear our environment.
# start by clearing your console. To do this hit Ctrl+l or go to Edit-->Clear Console
# clear your environment and plots by hitting the broom icon in both those cells
# reset our working directory
workdir <- '/scratch/Users/username/day6/'
```

becomes

```
####### Running featureCounts
####### Author: Taylor Jones
####### Here we will learn how to download a package, what metadata table is (and why it is important),
####### and run featureCounts, which counts reads over genes.

#########################################################################################################################
####### We will want to start fresh and clear our environment.
# start by clearing your console. To do this hit Ctrl+l or go to Edit-->Clear Console
# clear your environment and plots by hitting the broom icon in both those cells
# reset our working directory
workdir <- '/scratch/Users/maallen3/day6/'
setwd(workdir)
getwd()

# Like before: most packages you can install with this syntax: install.packages("PACKAGE")  Such as this package: install
```

maryallen — maallen3@ip-172-31-38-192:/scratch/Users/maallen3/day6 — ssh maallen3@18.219.252.252 — 146×34

```
#!/bin/bash
#SBATCH --job-name=featurecounts                                  # Job name
#SBATCH --mail-type=ALL                          # Mail events (NONE, BEGIN, END, FAIL, ALL)
#SBATCH --mail-user=username@email.edu           # Where to send mail
#SBATCH --nodes=1                                       # Number of cores job will run on
#SBATCH --ntasks=4                                      # Number of CPU (processers, tasks)
#SBATCH --time=2:00:00                                  # Time limit hrs:min:sec
#SBATCH --partition compute                             # Job queue
#SBATCH --mem=4gb                                       # Memory limit
#SBATCH --output=/scratch/Users/username/eofiles/%x_%j.out
#SBATCH --error=/scratch/Users/username/eofiles/%x_%j.err


module load R/3.6.1
R_LIBS="/data/R-lib" #this is here becuase we don't all want to install feature counts on R in our own home directorys. BIT installed it here
```

becomes

```
#!/bin/bash
#SBATCH --job-name=featurecounts                                  # Job name
#SBATCH --mail-type=ALL                          # Mail events (NONE, BEGIN, END, FAIL, ALL)
#SBATCH --mail-user=allenma@colorado.edu         # Where to send mail
#SBATCH --nodes=1                                       # Number of cores job will run on
#SBATCH --ntasks=4                                      # Number of CPU (processers, tasks)
#SBATCH --time=2:00:00                                  # Time limit hrs:min:sec
#SBATCH --partition compute                             # Job queue
#SBATCH --mem=4gb                                       # Memory limit
#SBATCH --output=/scratch/Users/maallen3/eofiles/%x_%j.out
#SBATCH --error=/scratch/Users/maallen3/eofiles/%x_%j.err


module load R/3.6.1
R_LIBS="/data/R-lib" #this is here becuase we don't all want to install feature counts on R in our own home di

################### SET VARIABLES #####################################

FEATURECOUNTS=featureCounts.R

#######################################################################
################### PRINT JOB INFO ####################################
```

Step 3) Submit the sbatch script to the queue
The sbatch script runs the R script… how?

```
[maallen3@ip-172-31-38-192 day6]$ sbatch sr_featcounts.sbatch
Submitted batch job 1293
```

Step 4) If it works you will end up with a directory named counts in your working directory. In the counts directory you will end up with two files. One has the count data. One has information about the feature counts run.

```
[maallen3@ip-172-31-38-192 day6]$ ls
counts  featureCounts.R  sr_featcounts.sbatch
[maallen3@ip-172-31-38-192 day6]$ cd counts/
[maallen3@ip-172-31-38-192 counts]$ pwd
/scratch/Users/maallen3/day6/counts
[maallen3@ip-172-31-38-192 counts]$
```

```
[maallen3@ip-172-31-38-192 counts]$ more featureCounts_gene_rnaseq.txt
fc.annotation...c..GeneID...    chr21Eric.repA.RNA.sorted.bam   chr21Ethan.repA.
RNA.sorted.bam
LOC102724184    18      22
LOC102723996    676     1269
LOC102724023    825     1609
LOC102724132    10      31
LOC101928576    51      59
LOC102724159    369     866
LOC102724200    608     1027
LOC102724219    0       0
LOC102724334    4       0
LOC102724354    0       2
LOC102724385    1       0
LOC102724370    0       1
LOC102724398    0       0
LOC102724411    1       0
```

```
[maallen3@ip-172-31-38-192 counts]$ ls
featureCounts_gene_rnaseq.notes.txt  featureCounts_gene_rnaseq.txt
[maallen3@ip-172-31-38-192 counts]$ more featureCounts_gene_rnaseq.notes.txt

        ==========     _____ _    _ ____  _____  ______          _____
        =====         / ____| |  | |  _ \|  __ \|  ____|   /\   |  __ \
          =====      | (___ | |  | | |_) | |__) | |__     /  \  | |  | |
            ====      \___ \| |  | |  _ <|  _  /|  __|   / /\ \ | |  | |
              ====    ____) | |__| | |_) | | \ \| |____ / ____ \| |__| |
        ==========   |_____/ \____/|____/|_|  \_\______/_/    \_\_____/
       Rsubread 2.0.1

//========================== featureCounts setting ===========================\\
||                                                                            ||
||             Input files : 2 BAM files                                      ||
||                           o chr21Eric_repA.RNA.sorted.bam                  ||
||                           o chr21Ethan_repA.RNA.sorted.bam                 ||
||                                                                            ||
||              Annotation : chr21_genes.gtf (GTF)                            ||
```