

# STAR Alignment, TSCC

BMS Bootcamp 2019

# Schedule

- Recap yesterday
- Fastqc
- TSCC job submission
- Reading documentation
- Generate STAR alignment
- STAR mapping of our data

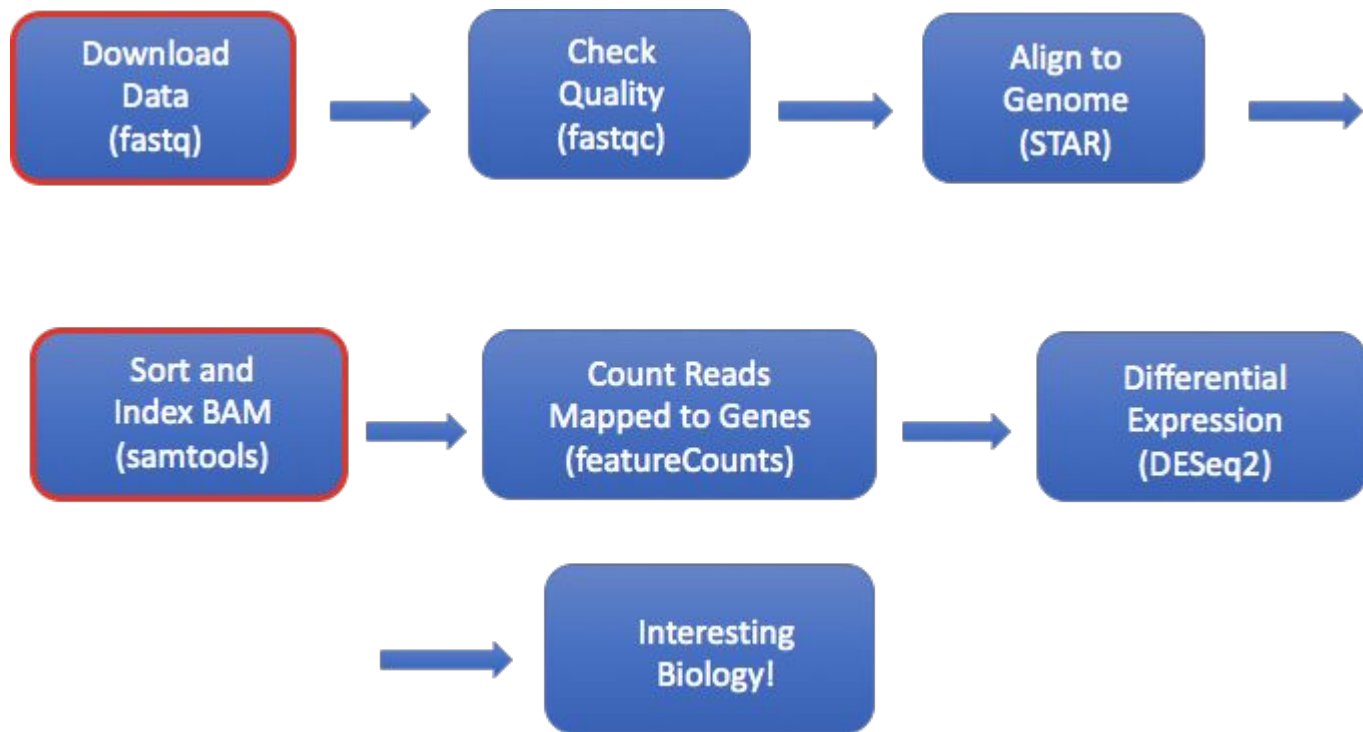
# Objectives

Be able to submit a job to the cluster

Be able to read documentation

Download genome alignment and genome annotation

# RNA-Seq Pipeline



# Submitting a Job to TSCC

- For tasks that require large amounts of processing power, you can submit a job to the TSCC computing cluster
- Submitting a job allows the user to request the specific number of compute nodes and processors per node needed for you job
- The submitted job goes into the designated queue and will be run when the requested number of processors becomes available

# TSCC Job Queues

- **TORQUE Resource Manager manages job queues**
  - Queues include:
    - hotel: supports all non-contributors to TSCC; this is the only node we are eligible to use
    - home: clusters reserved for members of particular group; purchased
    - condo: will allow contributors to run on nodes greater than those that have already been purchased; 8-hour time limit
    - glean: allows to run jobs, free-of-charge, on available idle nodes within condo; jobs will be terminated if idle nodes are requested

One or more job may be submitted simultaneously

# Submitting Jobs to TSCC using shell scripts

- Designed to be run on a Unix shell, command-line interpreter
- First line describes language to be used:
- We will submit jobs using bash scripting:
- `#!/bin/bash`
- PBS flags (`#PBS`) will be used to describe parameters of your job submission

```
#!/bin/csh
#PBS -q <queue name>
#PBS -N <job name>
#PBS -l nodes=10:ppn=2
#PBS -l walltime=0:50:00
#PBS -o <output file>
#PBS -e <error file>
#PBS -V
#PBS -M <email address list>
#PBS -m abe
#PBS -A <account>
```

# Checking the status of a submitted job

- `qstat -u <username>`

- Example: `qstat -u ecwheelee`

Q – job is in the queue to be run

R – Job is running

C – Job is complete

To delete a job:

`qdel <jobid#>`

You can find the jobid# from `qstat`

- Note – After your job has completed, the “C” will only be there for a few minutes. If the job has disappeared completely then it has finished running and you can find information about how it ran in your output files.



# Submitting an Interactive Job

- If you want to be able to interact with the command line while taking advantage of the compute power available on TSCC, you can submit an interactive job
- The job will be open for as long as the walltime specifies, and you will be operating on the number of nodes and processors requested

Example:

```
qsub -l -q hotel -l nodes=1:ppn=4 -l walltime=01:00:00
```

# STAR

Documentation <https://github.com/alexdobin/STAR/blob/master/doc/STARmanual.pdf>

We will use STAR to generate a genome map and also to align out fastq files to the genome build we made

# How do you align reads to a genome?

- 1) What is the sequence of our chromosomes - Reference Genome - Fasta file
- 2) What do we know about where genes are located on the chromosomes - Genome Annotation - GTF file