Exercises for RNA-seq: Alignment & Counting

EXERCISE 1

Run `ls` on the current directory. We need to clean things up before they get out of hand.

- Remove the trimmed file you just created
- Make a directory called "QC"
- Move all the fastqc output directories and zip files into the QC directory (hint, you can do this all with a wildcard and a single `mv` command)
- Use `find ... | parallel ...` to run `wc` on all the fastq files in parallel

EXERCISE 2

First, look around at the results, then delete those files when you're done:

- Go into the output directory that was just created.
- Look at the `align_summary.txt` file that was created (**don't try to `cat` the .bam files in that directory!**)
- When you're done, remove the entire directory containing that test output, and remove the test10k.fastq file you created.

Next, Using `find` and `parallel --dry-run`, try to construct the command that would run all the samples through tophat in parallel. Using the `-o` option, make the output for each run be a directory with `_tophat` appended to it. The command generated should look like this:

```
tophat --no-coverage-search -o trimmed_ctl1.fastq_tophat chr4 trimmed_ctl1.fastq
tophat --no-coverage-search -o trimmed_ctl2.fastq_tophat chr4 trimmed_ctl2.fastq
tophat --no-coverage-search -o trimmed_ctl3.fastq_tophat chr4 trimmed_ctl3.fastq
tophat --no-coverage-search -o trimmed_uvb1.fastq_tophat chr4 trimmed_uvb1.fastq
tophat --no-coverage-search -o trimmed_uvb2.fastq_tophat chr4 trimmed_uvb2.fastq
tophat --no-coverage-search -o trimmed_uvb3.fastq_tophat chr4 trimmed_uvb3.fastq
```

But, don't launch the jobs just yet

EXERCISE 3

- When that's done, from the main data directory, look at all the `align_summary.txt` files with one command.
- Using `grep`, pull out the line that shows you the number of mapped reads.