

Lab 16

Q1. How would you check that these files with extension ‘.fastq’ actually look like what we expect for a FASTQ file? You could try printing the first few lines to the shell standard output:

Q1A. `head SRR2156848_1.fastq`

Q2. How could you check the number of sequences in each file?

Q2A. `grep -c "SRR2156848_1" SRR2156848_1.fastq`

Q3. Check your answer with the bottom of the file using the tail command and also check the matching mate pair FASTQ file. Do these numbers match? If so why or why not?

Q3A. These numbers match because the total number of sequences should equal the number of the final sequence.

Q4. Can you run kallisto to print out its citation information?

Q4A. `kallisto 0.44.0`

Usage: `kallisto [arguments] ..`

Where can be one of:

<code>index</code>	Builds a kallisto index
<code>quant</code>	Runs the quantification algorithm
<code>pseudo</code>	Runs the pseudoalignment step
<code>h5dump</code>	Converts HDF5-formatted results to plaintext
<code>inspect</code>	Inspects and gives information about an index
<code>version</code>	Prints version information
<code>cite</code>	Prints citation information

Running kallisto without arguments prints usage information for

Q5. Have a look at the TSV format versions of these files to understand their structure. What do you notice about these files contents?

```
library(tximport)
folders <- dir(pattern="SRR21568*")
samples <- sub("_quant", "", folders)
files <- file.path( folders, "abundance.h5" )
names(files) <- samples
```