Zebrafish Dataset Practical 1

Before you start, make sure you’ve read the document that describes the zebrafish dataset we’re using in this practical. And make sure you’ve put the four required files (Amp.counts.tsv, Amp.samples.tsv, Oxy.counts.tsv and Oxy.samples.tsv) in your home directory.

To begin, here are a couple of exercises that require using the command line in Terminal:

1. Using the awk and wc commands, work out how many genes are significantly differentially expressed (adjusted p-value < 0.05) for the amphetamine and oxycodone treatments. How do these numbers change if you reduce the adjusted p-value threshold to 0.005 or even 0.0005?
2. Using awk, create two new files that just contain the subset of significantly differentially expressed genes (adjusted p-value < 0.05). Keep these two files as you’ll need them later in the week. Also, using cut, create two new files that just contain the Ensembl IDs of the significantly differentially expressed genes. Again, keep these two files for later.