1. Take this file, sort by the “FDR” column (retaining the header), only include the “Gene”, “logFC”, “FDR” columns, print the columns neatly (i.e. make them aligned) and redirect the output to a file called “newRNAseq1.txt”

Command: cut -f1,11,13 rnaseq.txt > newRNAseq.txt

(head -n 1 newRNAseq.txt && tail -n +2 newRNAseq.txt | sort -k 3) > newRNAseq1.txt

cat newRNAseq1.txt

Explain: Generally, it takes 2 steps to complete this task. Firstly, “cut -f1,11,13 rnaseq.txt > newRNAseq.txt” save the column 1, 11,13, which are the “Gene”, “logFC”, “FDR” columns respectively, and redirect to file ‘newRNAseq.txt’ by ‘>’. Second, sort the new file by column 3 but keep the head. The output was redirected to newRNAseq1.txt.

Reference:

https://unix.stackexchange.com/questions/35369/how-to-define-tab-delimiter-with-cut-in-bash

<http://www.briandolhansky.com/blog/2014/9/16/linux-snippets-sorting-a-file-ignoring-the-header>

2. Repeat except this time, sort by the “logFC” column. Redirect the output to a file called “newRNAseq2.txt”

Command: (head -n 1 newRNAseq.txt && tail -n +2 newRNAseq.txt | sort -k 2) > newRNAseq2.txt

cat newRNAseq2.txt

Explain: Here, in order to sort by the “logFC” column, we just need to change ‘-k 3’ to ‘-k 2’, and redirected it to “newRNAseq2.txt” .

3. Show what command line expression you could use to show the gene name, log fold change (logFC) and false discovery rate (FDR) for all the “Cxcr” or “Ccr” genes, sorted by FDR.

Command: grep -E “^C\*cr” newRNAseq1.txt > screen-newRNAseq1.txt

Explain: Since I have finished the cut and sort, and save it in the file newRNAseq1.txt, I just need to find the line that contain ‘Cxcr’ or ‘Ccr’. Here, I use the grep function with wildcard.

Reference: <http://swcarpentry.github.io/shell-novice/07-find/index.html>

4. Make an input file for Ingenuity Pathway Analysis. Such a file has three (tab-separated) columns (and no header row): gene name, log fold change, FDR, but restricted to only outputting genes where the FDR value is below a cutoff (let’s use 0.01). Do this and redirect the output to a file called “IPAinput.txt”

Sorry I can’t finish this task even if I search the resource online.

I do not know how to make the command into a separate file and make it run independently as the command.