Summary Report

**Introduction**

* Study investigates transcriptional differences between M. musculus (mouse), specifically the CD4 and CD8 T cells at conditions
* Sample cells
  + Control CD4 Day 0
  + Control CD8 Day 0
  + CD4 Day 7 FLICA Negative
  + CD8 Day 7 FLICA Negative
  + CD8 Day 7 FLICA Positive
* Understanding transactional differences between the populations in relation to activation and caspase activity, indicated by FLICA, can be beneficial
* Analysis will identify any differentially expressed genes (DEGs) and pathways that are affected by the DEGs

**Methods**

* RNA seq data was processed by indexing, then reads were quantified using Kallisto
* Kallisto used the Mus musculus reference transcriptome to produce an alignment and estimate transcript abundances
* The expression data was put through differential gene expression analysis by NetworkAnalyst.ca
* Preprocessing included quality check, filtering, and normalization.
* Analysis was performed with the Limma package

The number of significant genes remaining after filtration was 8941

The differential expression results after quality checking and data filtering/normalization with logCPM transformation

A graph with rectangular objects

Description automatically generated with medium confidence

Fig 1 qc\_norm\_boxplot\_0\_dpi72

Fig 1 is a boxplot that displays the distribution of normalized gene expression values across the 15 samples. The median of all of the samples seem to stand right below 6, meaning that the samples are similarly comparable in terms of expression.

A graph with colored lines

Description automatically generated

Fig 2 qc\_norm\_density\_0\_dpi72

The density plot shows that the curves of the sample distributions overlap largely well, which indicates that they are similar overall expression distributions. The distribution is skewed right and multimodal, with significant peaks at around value 4 and 6.

A graph with different colored dots and numbers

Description automatically generated

Fig 3 qc\_norm\_pca\_0\_dpi72

This PCA plot above clusters the samples based on similarities. Since there are 3 of each sample, there will be 3 dots clustered. Principal component 1 tries to explain 26.6% of the variance while the principal component 2 explains 9.7% of the variance, together capturing 36.3% of the total variation. Interesting to note that CD8 and CD4 control samples share the most similar identities, shown by how close they are grouped together. The FLICA negative samples are also grouped relatively close together, although more apart than the control samples. However, the FLICA positive CD8 sample seems to be the furthest separated from the other groups, showing that the effect of activation state seems to make a large difference.

A screen shot of a graph

Description automatically generated

Fig 4 Volcano plot of enrichment analysis between CD4 and CD8 control groups

The plot has an x-axis (log2(FC)) and y-axis (-log10(p-value)). The grey dots at the bottom are unsignificant while the blue dots represent downregulated, or reduced transcription/translation, genes and red dots represent some of the upregulated, or increased transcription/translation, genes. There is an extreme outlier gene Cd8a that is massively downregulated from the rest of the genes (P-val: 2.2787e-8, Log-FC: -7.3820).