

PHS 675 Special Topics: Analysis of RNA-Sequencing Data
Homework 1: Differential Expression. Model Solution.

Methods:

Male TALLYHO ("TH"; n=7) and C57BL/6 ("B6"; n=7) mice were fed a high-fat diet from weaning until 20 weeks of age. At that time, the mice were sacrificed and the pancreas extracted from each mouse. RNA was extracted and sequenced on an Illumina NextSeq 2000 sequencer using a paired-end 50bp sequencing strategy and standard sequencing protocols.

Data were quality controlled and then analyzed to compare expression between TH mice and B6 mice using DESeq2 (Love et al., 2014) version 1.48.2 running on R version 4.5.3.

Results:

Quality assessment of the fold change computations by MA-plot (figure 1) and PCA plot (figure 2) showed expected distributions of fold changes and good separation of samples by strain, respectively.

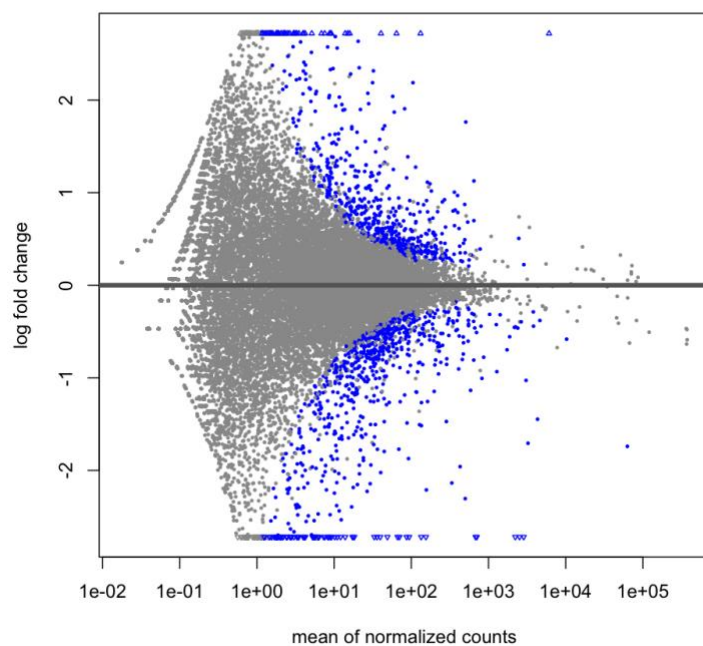


Figure 1. MA plot showing log fold change against overall expression for each gene.

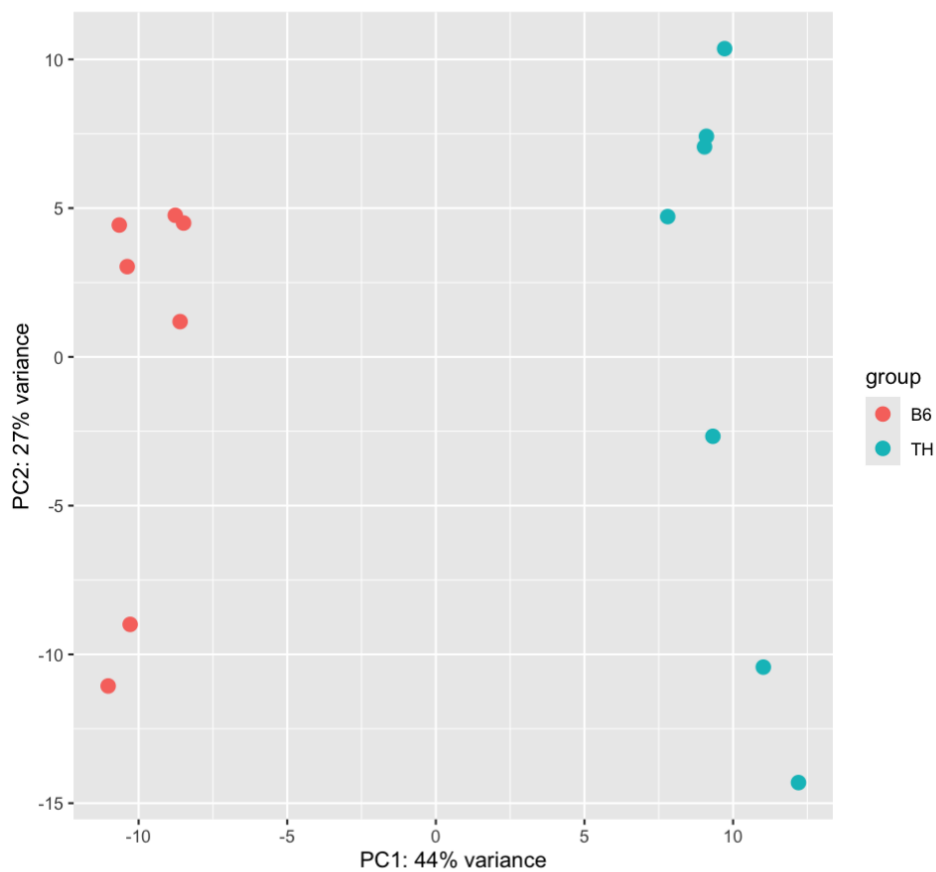


Figure 2. PCA plot of samples.

Using a threshold for statistical significance of a Benjamini-Hochberg adjusted p-value of 0.1, 605 genes were shown to have increased expression in the TH mice compared to the B6 mice, and 737 genes were shown to have decreased expression in the TH mice compared to the B6 mice (Figure 3). The top 20 genes by absolute value of fold change are shown in table 1 and figure 4.

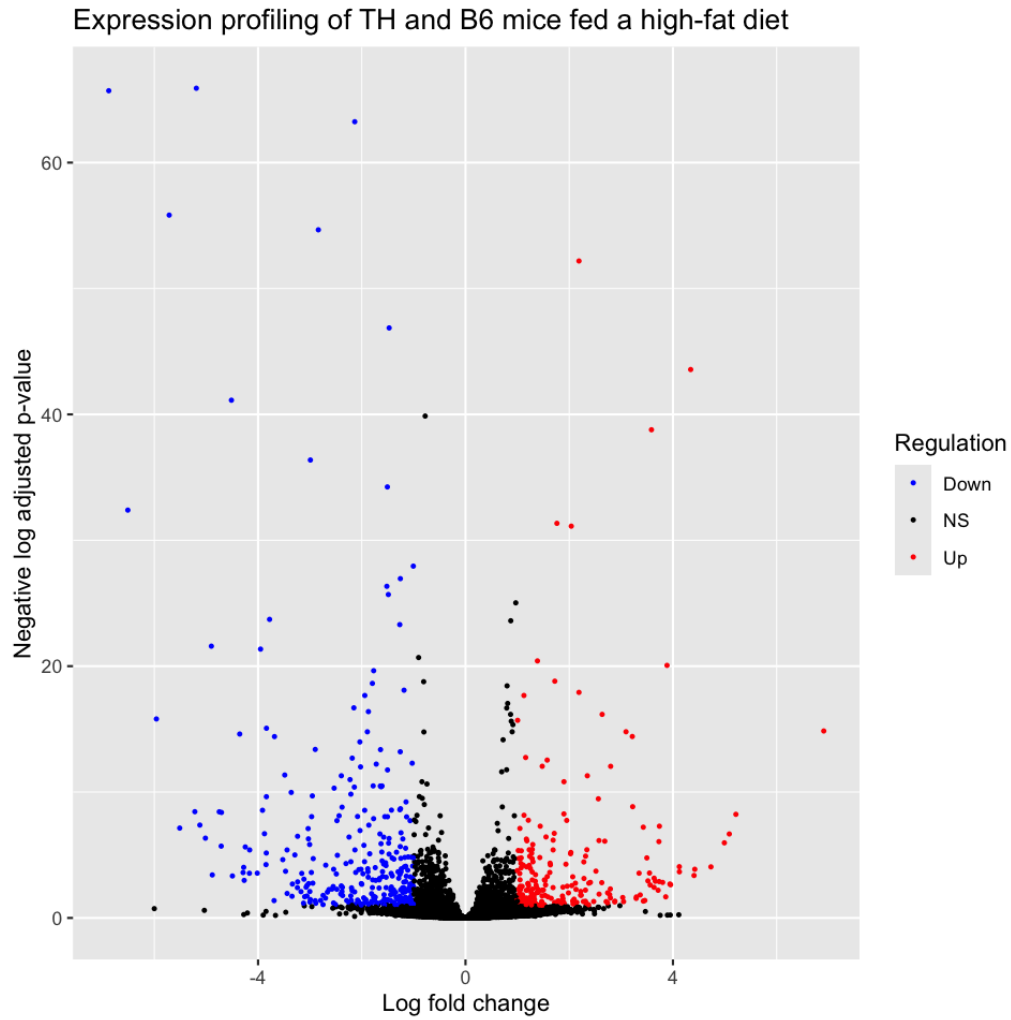
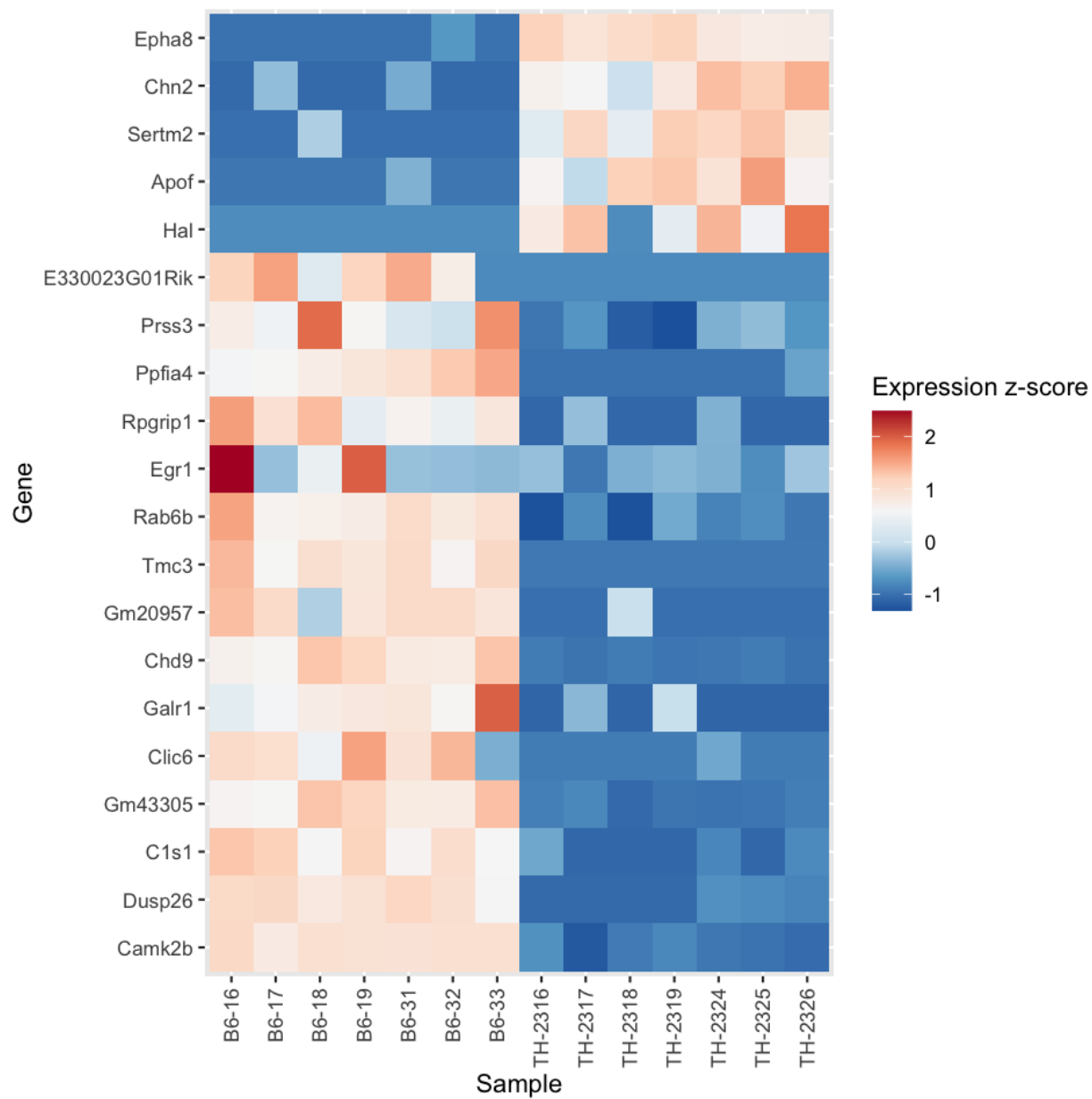


Figure 3. Volcano plot of expressed genes. Up- and down-regulated genes are determined by an adjusted p-value (false discovery rate) less than 0.1 and a log base 2 fold change of at least 1 in absolute value.

Gene	Log base 2 fold change	BH adjusted p-value
Camk2b	-6.88	1.95E-66
Dusp26	-6.51	4.02E-33
C1s1	-5.96	1.56E-16
Gm43305	-5.71	1.48E-56
Clic6	-5.51	7.29E-08
Galr1	-5.22	3.63E-09
Chd9	-5.19	1.23E-66
Gm20957	-5.12	4.20E-08
Tmc3	-5.02	4.59E-07
Rab6b	-4.90	2.58E-22
Egr1	-4.88	3.83E-04
Rpgrip1	-4.75	3.71E-09
Ppfia4	-4.71	1.98E-06
Prss3	-4.71	4.17E-09
E330023G01Rik	-4.49	4.64E-04
Hal	4.73	8.76E-05
Apof	4.99	1.08E-06
Sertm2	5.09	2.14E-07
Chn2	5.21	5.88E-09
Epha8	6.91	1.41E-15

Table 1. Top 20 genes by absolute value of fold change.



Heatmap of the top 20 differentially expressed genes (determined by selecting for false discovery rate less than 0.1 and ranked by absolute value log base 2 fold change).

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
<https://doi.org/10.1186/s13059-014-0550-8>