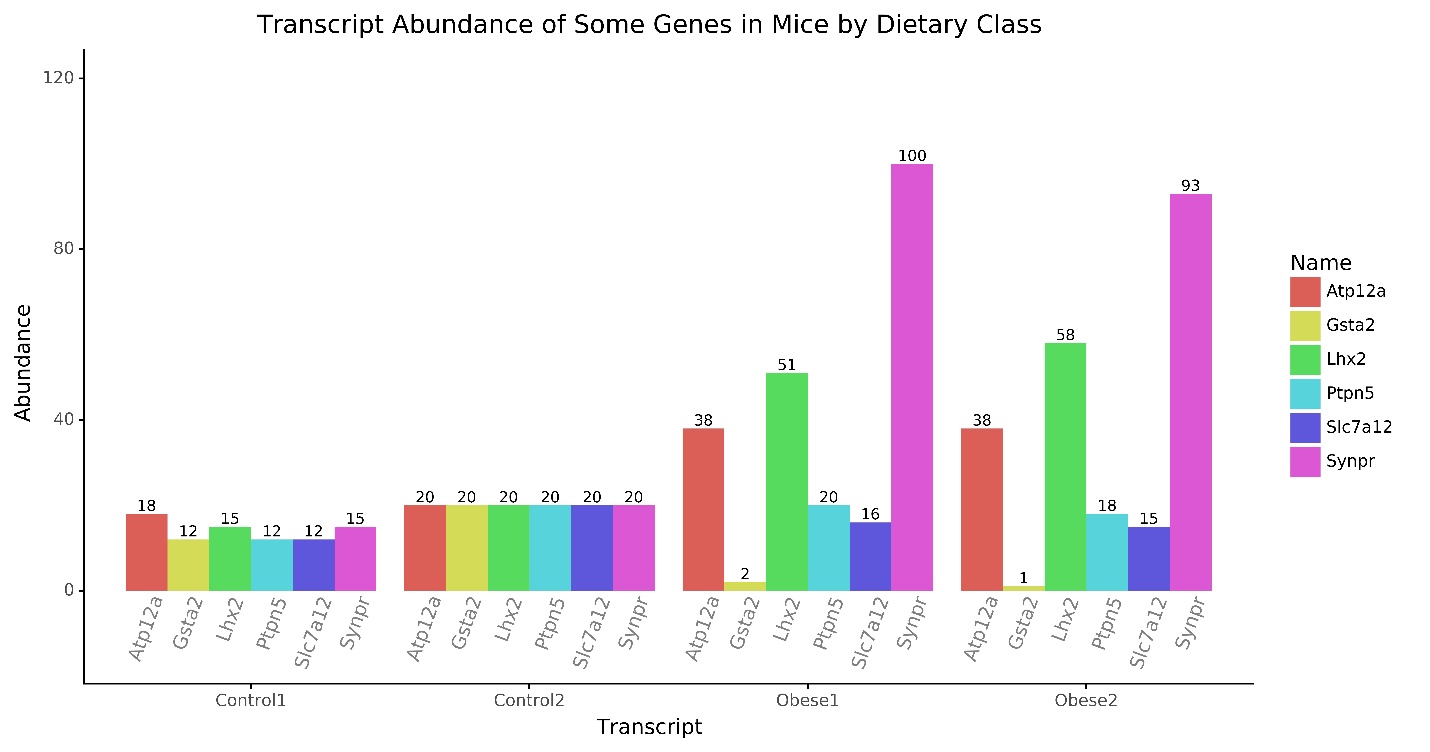
Biocomputing Group Project Data comparison and Short Answers

Results Comparison



The results provided here corroborate the Kuhns and Pluznick study. The original study’s results are reported as fold changes with fold decreases in Slc7a12 and Gsta2 and fold increases in Atp12a, Lhx2, ptpn5, and Synpr in High fat mice versus controls. We show decreases in the total abundance of the same transcripts in which they reported fold decreases and increases in the total abundance of the same transcripts in which they report fold increases.

Short Answers

1. Effects of Blast Type

|  |  |  |  |
| --- | --- | --- | --- |
|  | megablast (default)  #Hits mouse only | Discontiguous megablast #Hits mouse only | Blastn #Hits mouse only |
| Gsta2 | 42 | 54 | 101 |
| Ptpn5 | 27 | 118 | 370 |

When restricted to the mouse database, there appears to be just as many dissimilar sequences for Gsta2 in the database as highly similar sequences. Of course, a blast of somewhat similar sequences yields more hits than the standard blast for highly similar sequences.

Comparing Ptpn5 blast results amongst the three modes suggests there’s quite a bit of divergence in the sequence. Again, using a blast for somewhat similar sequences increases the number of hits.

In general, the type of blast search to be used is dependent on the gene and what you are trying to do with it. Here, the goal is to look at transcript abundance in a single mouse species for a specific experiment, so the blast that yields the most similar sequences is the most useful. The other modes would likely be more useful for phylogenetic comparison.

2. How does building a hmm protein model based on Gsta2 and Ptpn5 from a distantly related mammal, a platypus, influence search quality using RNAseq data from a mouse?

Building a HMM protein model based on platypus Gsta2 (glutathione S-transferase) and Ptpn5 (tyrosine phosphatase) sequences negatively impacted the number and quality of hits when searching the RNAseq files. Comparing the search hit tables of the mouse HMMs (HMMsearchHits.txt) and the platypus HMMs (HMMsearchHits2.txt), the mouse HMMs generated on average more hits than the platypus HMMs. Each mouse HMM generated at least 1 hit while the platypus tyrosine phosphatase HMM had no hits in any of the 4 RNAseq files. The quality of the hits also decreased when using the platypus HMMs, which was reflected in the generated e-values. E-values describe the background noise of the sequence and correspond to the number of hits expected from random chance. Therefore, a low e-value would mean a higher identity score between sequences. The mouse HMMs had much lower e-values than the platypus HMMs. For example, the mouse Gsta2 (transcript 1) had an e-value of 1e-158 for Obese2 while the platypus Gsta2 had an e-value of 9.6e-66. The e-value of mouse Ptpn5 could not be compared to the e-value of platypus Ptpn5 because the platypus HMM generated no hits. Therefore, the number and quality of hits are not as good for an HMM built using a distantly-related mammal. This could also suggest that the sequences for these proteins are nonconserved and species-specific.