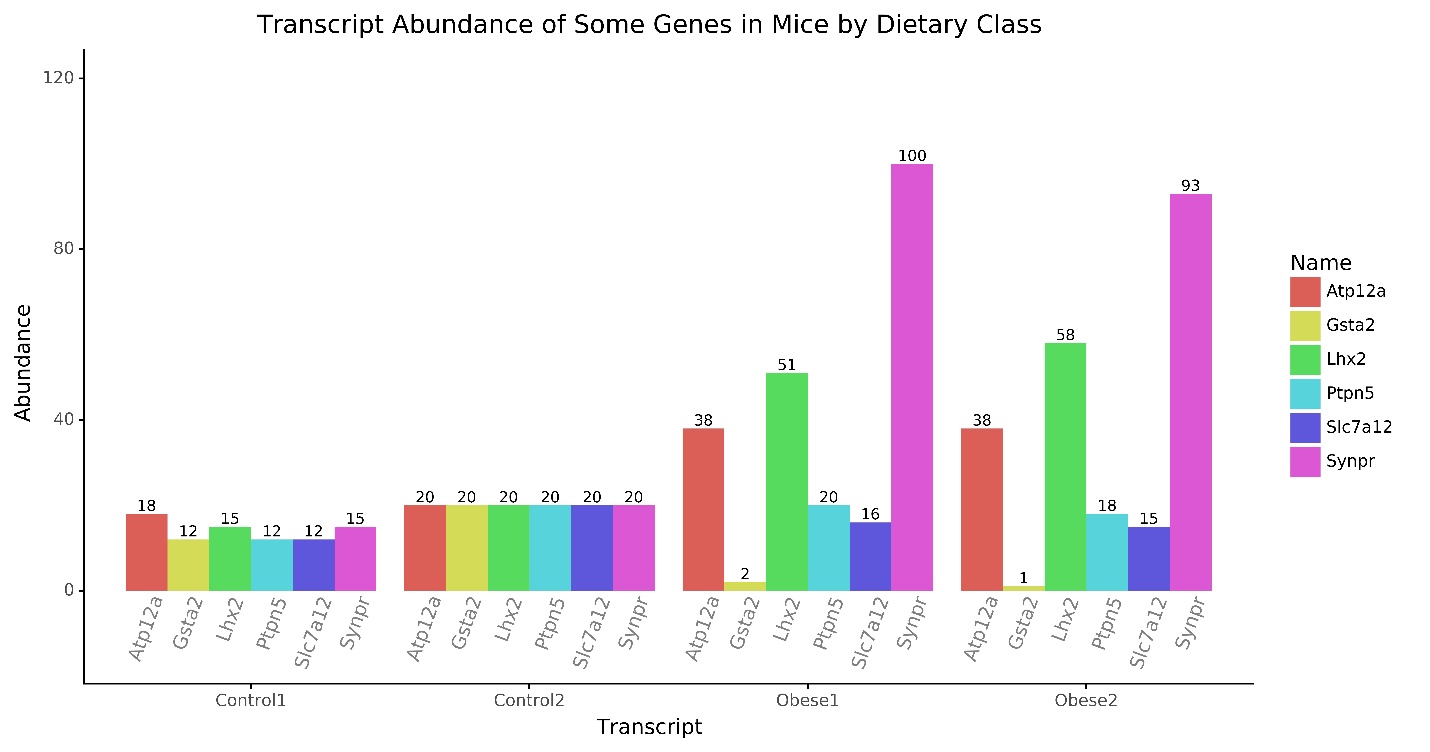
Biocomputing Group Project Data comparison and Short Answers

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 they reported fold decreases in and total abundance increases in the same transcripts they report fold increases in.

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?