

Stats 530
Hw 4
Mateusz Grobelny

1. To determine whether the distance d_j between cis-eQTL and the transcripts of classes A and B differ, I would use the Kolmogorov-Smirnov test.

This test would allow me to measure the distance between the cumulative distribution of d_j between class A vs B. Such that the KS test statistic T :

$$T = \sqrt{\frac{LL'}{L + L'}} \sup_{0 \leq t \leq 1} |\hat{F}_1(t) - \hat{F}_2(t)|,$$

Where L and L' are the sizes of the two distributions and \hat{F}_1 and \hat{F}_2 are the two distributions of d_j .

2.

For the CRBL region exhibited the largest difference in expression pattern, the MEDU and WHMT regions clustered together suggesting similar expression patterns. The remaining 7 brain regions all clustered together and had very summary expression values for all markers. Compared to Fig 1a, the expression patterns are quite similar as CRBL clustered by itself and MEDU and WHMT also clustered together. However, for the remaining 7 regions the patterning was different as SNIG and THAL had clustered together with MEDU and WHMT in the Fig1a but not in the heatmap generated in the HW.

6.

Number of Unique transcripts:

Cluster 1 – 13

Main Functional Annotation Clusters: Ion transport, DNA Binding, membrane related
10 terms not used

Cluster 2 – 10

Main Functional Annotation Clusters: Ion binding and transport, metal binding, Signal Peptide
4 terms not used

Cluster 3 – 40

Main Functional Annotation Clusters: Axon related, Locomotory behavior, Lipo protein related, Nucleoside binding, microtubule processes, cytoskeleton, protein complex assembly, golgi apparatus, apoptosis related, protein kinase, ion binding, membrane related.
30 terms not used

