Julian Moon  
BE504

Differential Expression of the SLC12A Family in ES and NPC Cells

**Abstract**

GABA is a neurotransmitter that changes role between embryonic development and maturity in the brain of humans. Using RNA-sequencing data of ES cells and NPCs, the differential expression of a portion of the genes that are responsible for this change was observed. After adjusting the data and performing various testing methods as well as multiple testing correction, the study has found a total of 11 genes that had significant changes in means. Although the results agree with previous literature on specific genes that are known to be differentially expressed, the limitations of the testing methods, as well as the data itself, suggests more work is needed to fully understand the changes observed in all the other genes.

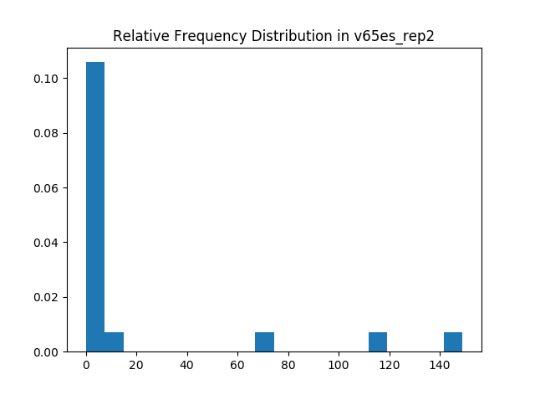
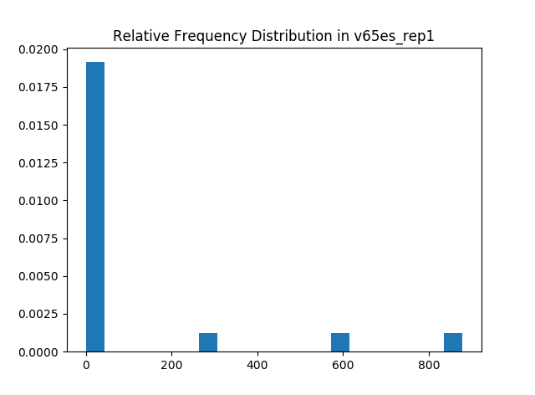
**Introduction**

GABA-A receptors are ligand-gated ion channels selective for chloride. During embryonic development, intracellular neuronal chloride concentrations are elevated, causing GABA to act as an excitatory neurotransmitter, but in a mature neuron, GABA acts as an inhibitory neurotransmitter,. This sequence is observed in a wide range of brain structures and animal species suggesting that it has been conserved throughout evolution. In embryonic cells, the GABAergic depolarization cells acts in synergy with N-methyl-D-aspartate (NMDA) receptor-mediated and voltage-gated calcium currents to enhance intracellular calcium exerting trophic effects on neuritic growth, migration and synapse formation. This process is mediated primarily by a developmentally regulated expression of the NKCC1 and KCC2 chloride importer and exporter. In an immature brain, an upregulation of NKCC1 and a downregulation of KCC2 causes the increase of intracellular chloride ion concentration. KCC2 and NKCC1 are coded by genes in the SLC12A transporter family. This solute carrier family consists mostly of K-Cl cotransporter, but also includes a bumetanide-sensitive Na-K-Cl cotransporter. Using RNA-seq data in embryonic stem cells and neural progenitor cells, the differential expression in these genes were explored.

**Hypotheses**

Since in an immature brain, there is a down regulation of KCC2 and an upregulation of NKCC1, the null hypothesis is that there is no difference in mean in counts for the genes that are responsible for KCC2 and NKCC1. The alternative hypothesis is that for ES cells, the mean for KCC2 will be lower and the mean for NKCC1 will be higher than NPC cells.

**Method**  
KCC2 and NKCC1 are primarily responsible for the difference in intracellular chloride concentration, but to expand the scope of the study, all the members in the solute carrier family was examined. There was a total of 19 genes that were responsible for the nine members of the SLC12A family. After filtering out five genes that were not expressed in any of the replicates, histograms and scatterplots were employed to determine necessary adjustments for the data.



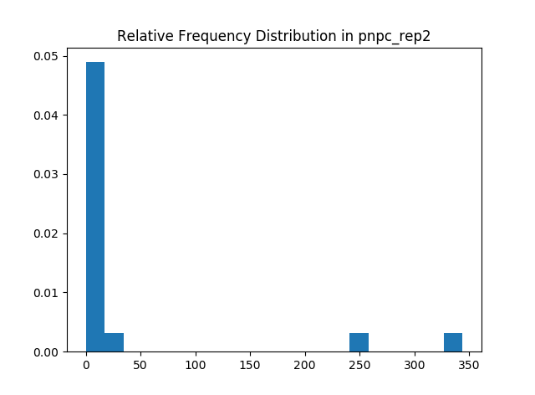
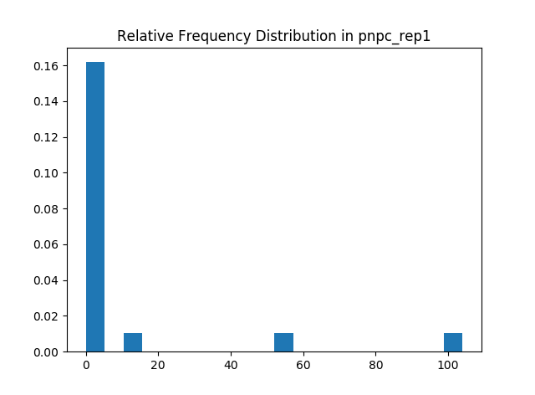


Figure 1. As expected, even after zero-filtering, the untransformed data is extremely skewed to the right. In addition, there are too many empty bins for the counts, making the data difficult to use for statistical testing.

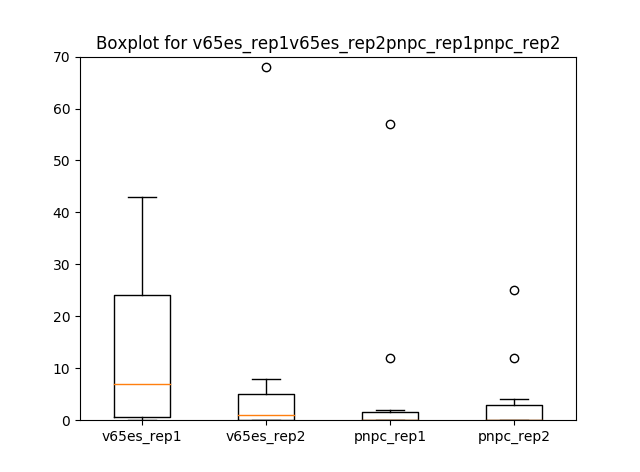
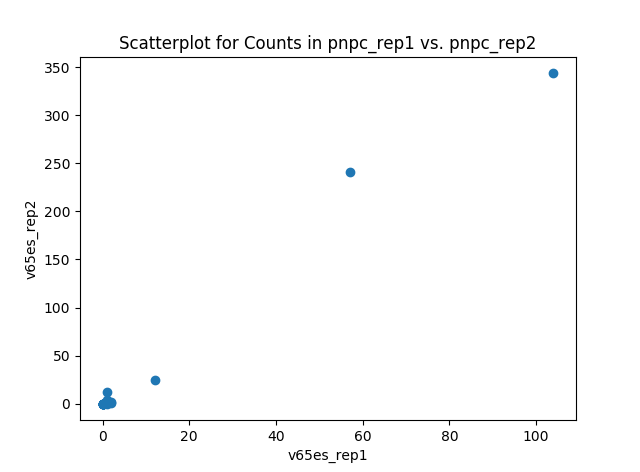
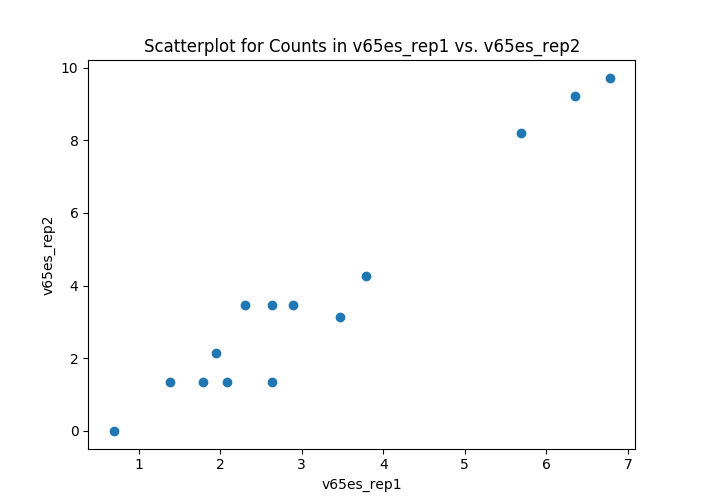


Figure 2. The box plot for all four replicates indicate that most of the values lie close to zero with some outliers.

The log transformation and normalization using the median-of-ratios method was selected for optimal results. TPM would have been useful, but was not chosen because only five out of the fourteen filtered genes would have been available for analysis due to lack of gene information.



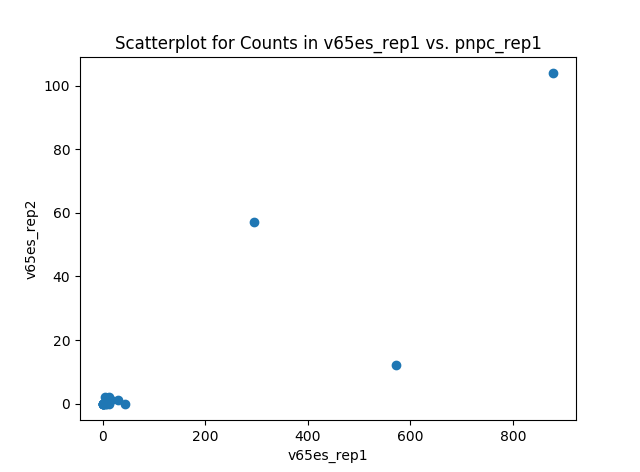
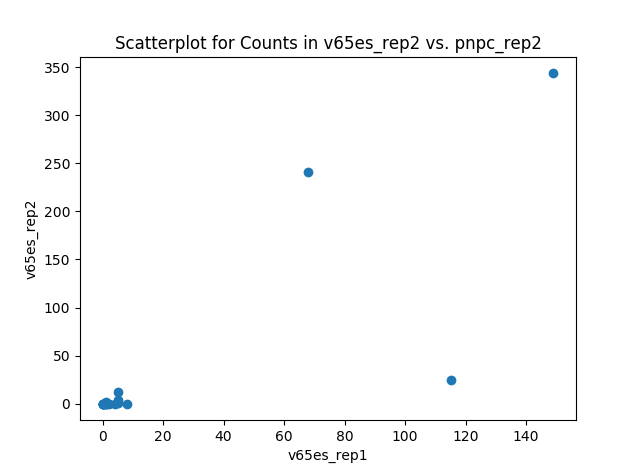
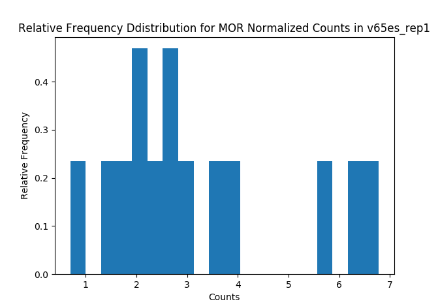
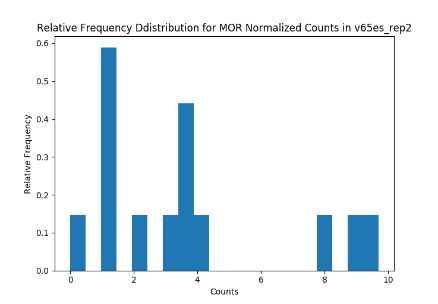


Figure 3. The transformed scatterplot for all combinations of the replicates are shown. Within ES and NPCs, the linear trend indicates that the choice for normalization was an appropriate one.



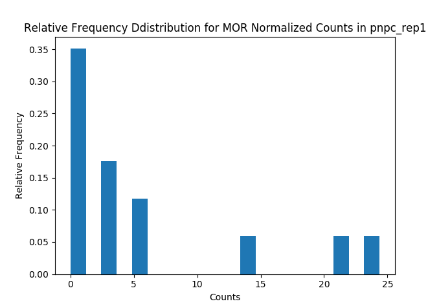
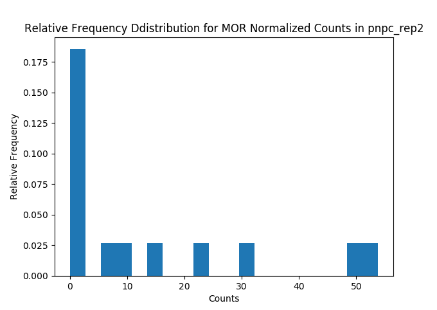
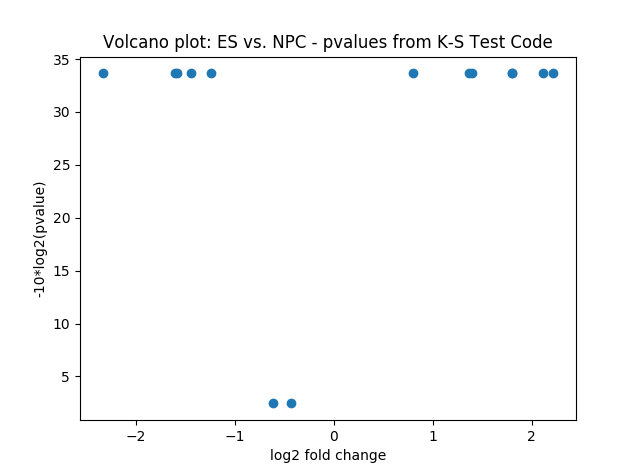
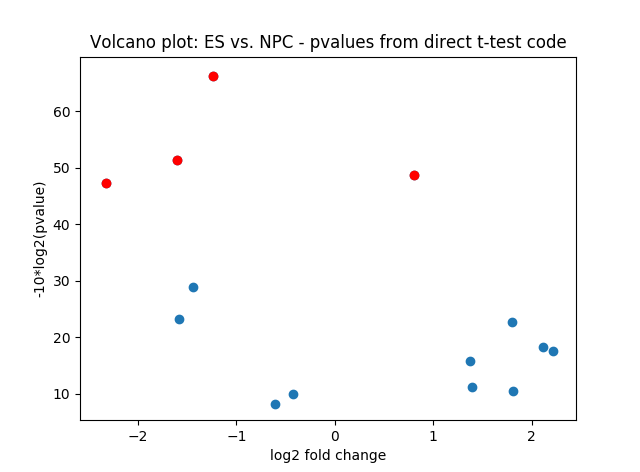


Figure 4. The histograms for all four replicates after normalization by median-of-ratio still shows strong positive skew. However, TPM was unavailable as a method, and the transformed distribution still has a preferable shape to the original one.

After transformation, p-values were obtained using Welch’s t-test, Fisher’s exact test, Kolmogorov-Smirnov and Poisson log-likelihood ratio test. Finally, multiple testing correction using the Benjamini-Hochberg method was performed to adjust the p-values for more accurate analysis.



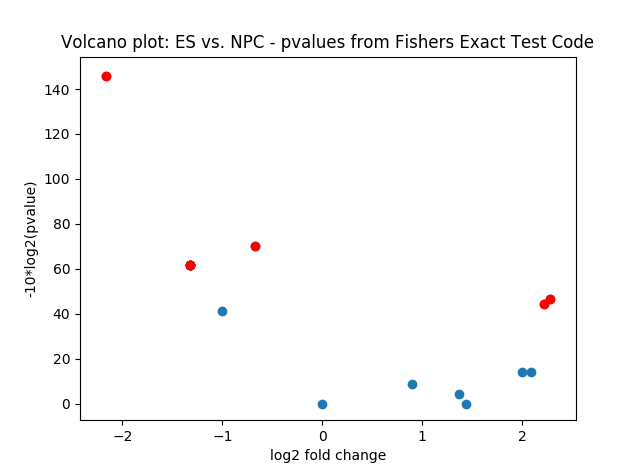
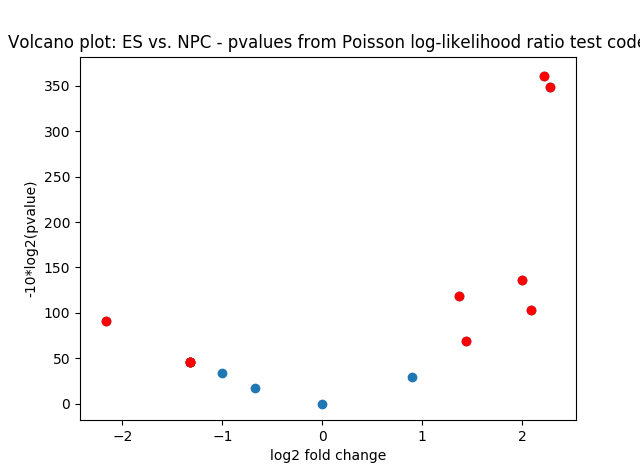


Figure 5. The volcano plots show the significant p-values (before multiple testing correction) in red, and their corresponding log2 fold change. Noting that a significant p-value should have a relatively large fold change, the LRT appears to perform the best, then Fisher and the t-test. K-S test did not produce any significant results.

**Results**

At the alpha level of 0.05, Fisher’s exact test and Poisson LRT yielded 5 and 7 significant p-values respectively after multiple testing correction. Five genes that code for SLC12A6, SLC12A7 and SLC12A8 were upregulated in ES cells according to Fisher’s exact test. Likewise, seven genes that code for SLC12A2 (NKCC1), SLC12A5 (KCC2), SLC12A6, SLC12A7, SLC12A8 and SLC12A9 were significantly changed in ES cells according to the LRT. Three of them were most likely downregulated with the treatment considering they had 0 counts for both NPC replicates, but four had relatively similar counts in all four samples, making it difficult to discern which direction gene expression took during the treatment.

The first thing to note about the results after multiple testing correction is the polarized results obtained through various testing methods. While Welch’s t-test and K-S test yielded no significant p-values, Fisher’s exact test and the log-likelihood ratio test yielded two almost completely opposite results. The possible deviation of the t-test could be explained by the fact that the counts data, with only four replicates, is nowhere close to satisfying the strict assumptions of normality required for the test. The K-S test observes the maximum difference between the cumulative distributions for the two types of cells, and because there are only two replicates for each sample, the test may not have produced reliable results. Fisher’s exact test uses 2x2 contingency tables to calculate p-values directly using a hypergeometric distribution, so it should have produced more accurate results. The LRT test makes the broad assumption that the counts follow a Poisson distribution, which is not accurate because the variance is much greater than the mean in the data. However, it is worth noting that the gene responsible for NKCC1 and one of the genes responsible for KCC2 were significantly upregulated according to the LRT. This produces slightly ambiguous results, as research has shown that in ES cells, while NKCC1 should be upregulated, KCC2 should be downregulated. Two genes code for KCC2, and one of them had zero counts for all four samples.

Another point to note is that among the significant genes, only one overlapped in both testing methods. This is most likely due to the fact that the two testing methods employ different mechanisms to calculate the p-value. However, with background information that can help us confirm the results, it is possible that LRT produced more accurate results in this study.

**Conclusions**

Regardless of the specific direction of regulation from ES cells to NPCs, both Fisher’s exact test and LRT show statistically significant changes in counts in a given portion of the genes that code for the SLC12A family. Specifically, genes that code for NKCC1 and KCC2 both show significant changes, which supports the claim that there is a difference in intracellular chloride concentration, which in turn illustrates the different role that GABA plays as a neurotransmitter in ES cells and in adult brain cells.

**References**

Y. Ben-Ari. *The GABA Excitatory/Inhibitory Developmental Sequence: A Personal Journey.* Neuroscience 279 (2014) 187-219