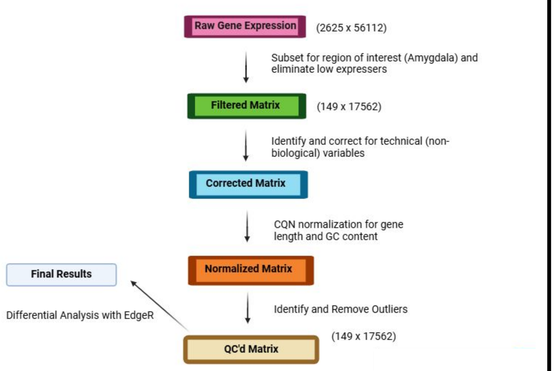
Sex Differences in Gene Expression in the Amygdala: Transcriptomic Insights into Neurodevelopmental and Neuropsychiatric Disorders

Jesus V. Velazquez, ,Ramin Ghoddousi, and Daniel Geschwind
Departments of Neurobiology and Psychology, UCLA, Los Angeles , CA 90024

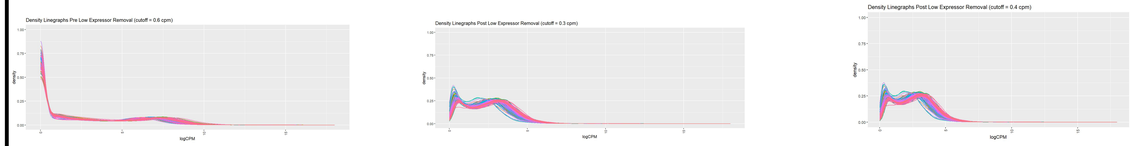
Introduction

Sex differences in the prevalence, onset, and phenotypic profiles of various neurodevelopmental and neuropsychiatric disorders. These differences suggest underlying biological mechanisms that contribute to sex-specific vulnerabilities and manifestations of these conditions. Current research efforts aim to elucidate the genetic, transcriptomic factors driving these sex differences. This study focuses specifically on the transcriptomic aspect, exploring gene expression variations between male and female brains. We utilized adult brain tissue samples from both sexes, with bulk RNA-sequencing data acquired from a well-characterized dataset. The data acquisition process involved selecting high-quality brain region samples, followed by rigorous preprocessing steps to ensure data accuracy and reliability. Quality control was meticulously performed to eliminate noise, outliers, and non-biological factors that could contribute to variability in the data. This cleaning process was essential to ensure that the subsequent analyses were focused on true biological differences rather than technical artifacts. The research aims to identify differentially expressed genes (DEGs) between sexes within a specific brain region. By comparing gene expression profiles between male and female brains, we seek to uncover sex-specific DEGs and investigate their potential associations with neurodevelopmental and neuropsychiatric disorders. To identify significantly expressed genes, we employed volcano plots, which allowed us to visualize and prioritize DEGs based on their statistical significance and fold-change.

Methods Summary

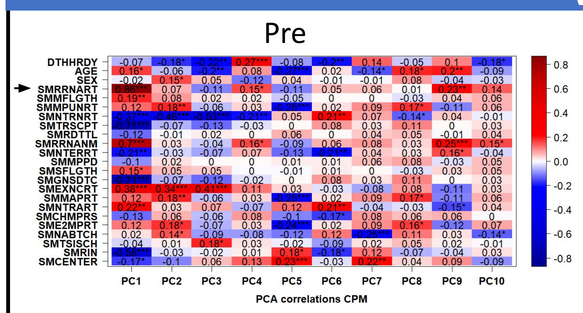


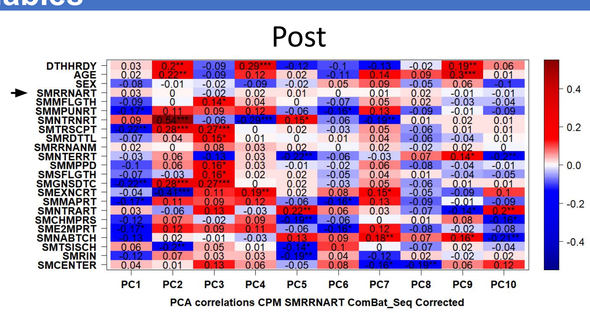
Low Expressor Removal: Cutoff Selection and Impact on Data Quality

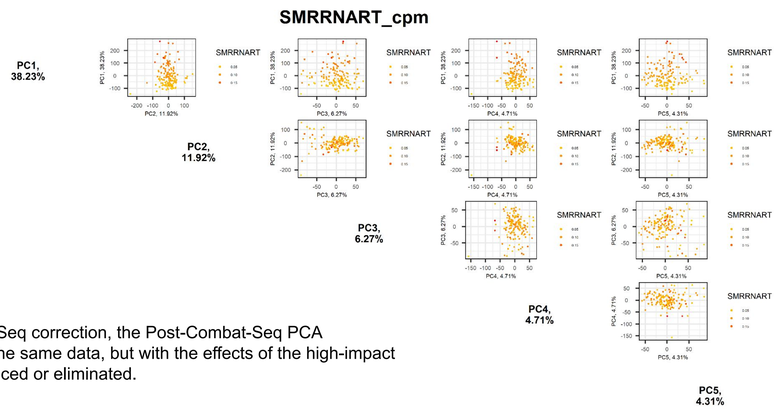


In the process of refining the dataset, a crucial step was the removal of low expressor genes, which are genes with minimal expression across samples. These low expressor genes can introduce noise into the analysis, potentially obscuring true biological signals. To mitigate this, I applied a cutoff based on counts per million (CPM). After evaluating multiple cutoff values, I selected a cutoff of **0.4 CPM**. This threshold was chosen because it effectively removed genes with consistently low expression across samples

PCA Correlation Plots: Pre and Post Combat-Seq Correction for Technical Variables



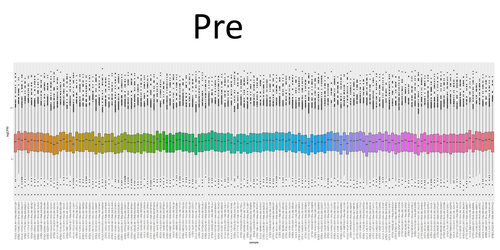


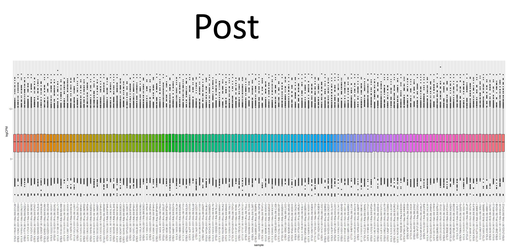


After applying Combat-Seq correction, the Post-Combat-Seq PCA correlation plot shows the same data, but with the effects of the high-impact technical variables reduced or eliminated.

The goal of Combat-Seq is to adjust for batch effects while preserving the biological variance of interest, such as differences between experimental groups. In the post-correction plot, the technical variables that previously dominated the variance should have a much-reduced impact, allowing biological factors (e.g., sex, age) to emerge more clearly in the principal components.

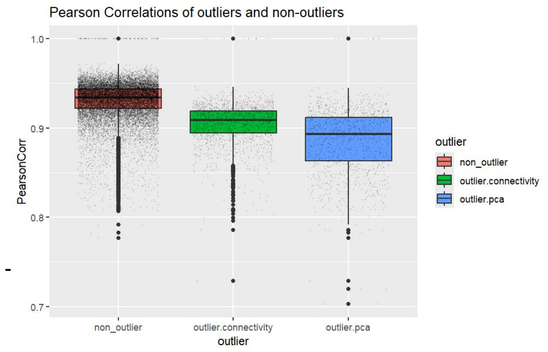
CQN Normalization: Comparison of Pre and Post Normalization Boxplots





After applying CQN, the post-normalization boxplots illustrate a significant reduction in the variability seen across samples. The boxplots are more consistent in their medians and ranges, indicating that the technical biases have been successfully corrected.

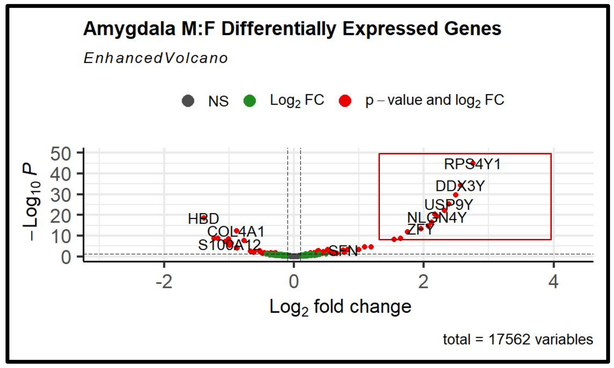
Outlier Detection and Removal: Boxplot with Detection Methods



Samples with low connectivity are potential outliers, as they do not align well with the typical expression patterns observed in other samples. The outlier connectivity plot visually represents these low-connectivity samples, making it easier to spot those that stand out.

By applying these outlier detection methods and visualizing the results through these plots, the data is refined by removing samples that could skew the analysis, ultimately leading to more accurate and reliable conclusions.

Differential Expression Results: Volcano Plot of Key Genes



Conclusion

- During the outlier detection process, we identified **20 outliers** in the dataset.

Following the outlier removal and differential expression analysis, several significant upregulated genes were identified. Among these, the following genes stand out due to their relevance to neurodevelopmental processes and potential links to Autism Spectrum Disorder (ASD):

- **RPS4Y1**:
  - Y-linked ribosomal protein involved in protein synthesis.
  - Has been implicated in sex-specific differences in brain function and development.
  - Some studies suggest a role in conditions like Turner syndrome, which can have neurodevelopmental aspects.
- **DDX3Y**:
  - Y-linked RNA helicase involved in RNA metabolism and regulation.
  - Homologous to DDX3X, which has been associated with intellectual disability and ASD.

We plan to extend this differential expression analysis to multiple other brain regions. By doing so, we aim to identify whether the upregulated genes observed in this study are consistently differentially expressed across various parts of the brain.

We will also quantify the overlap between the differentially expressed genes identified in this and future analyses with known ASD and other neurodevelopmental disorder risk genes.