**Distribution Agreement**

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Yutong Liu Date

Multi-tissue differential expression analyses between proxy case and control samples

By

Yutong Liu
Master of Science in Public Health

Department of Biostatistics and Bioinformatics

---

Steve Qin, PhD
Committee Chair

---

Yijuan Hu, PhD
Committee Member

# Multi-tissue differential expression analyses between proxy case and control samples

By

Yutong Liu

B.S.A., Zhejiang University, 2021

Thesis Committee Chair: Steve Qin, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health in Biostatistics
2023

**Abstract**

Multi-tissue differential expression analyses between proxy case and control samples

by Yutong Liu

**Background:** The Polygenic Risk Score (PRS) is a widely adopted approach to amalgamate information from multiple genetic locations to determine the risk of an individual developing a specific medical condition, including Alzheimer's disease (AD). AD is a heritable disease, with over 50 risk loci being identified with genome-wide significance. Computational methods, such as machine learning and artificial intelligence-driven meta-analysis, are being utilized to reveal more novel genes linked to AD, leveraging data from genome-wide association studies (GWAS).

**Objectives:** This study aims to explore the potential of polygenic risk score (PRS) in defining proxy cases and control samples and identify potential biomarkers based on the proxy case and control samples.

**Methods:** PRS scores are generated for each individual in Genotype-Tissue Expression (GTEx) database by weighting their genotype at each SNP by the corresponding log odds ratios provided in the PRS. Differential expression analysis is conducted between groups with upper 10% PRS and lower 10% PRS, and genes that exhibit significant up or down-regulation are identified (adjusted p-value $< 0.05$). The DisGeNET database is utilized to conduct chi-square tests to evaluate the extent of enrichment of AD-related genes that were identified through differential expression analysis in each tissue.

**Results:** PRS scores calculated for participants in GTEx are normally distributed, with a mean of 2.63 and a standard deviation of 0.33. The differential expression analysis across all tissues identifies 170 genes with significant up or down-regulation, with 154 in brain tissues and 18 in whole blood. Several highly significant genes are found related to AD, including *MT-TL2, POMC*, *HTR2C*, *CARTPT*, *KLHL7-AS1*, *COL24A1*, and *SOX14*. Chi-square tests revealed that AD-related genes are significantly enriched by differential expression analysis stratified by PRS in 12 of 14 tissues.

**Conclusions:** PRS is useful in defining proxy cases and control samples in AD. Analysis of 13 brain tissues reveals several genes with substantially different expression levels in individuals with higher versus lower PRS scores. PRS is effective in enriching AD-related genes across multiple tissues, while the inclusion of unrelated genes with high significance suggests the need for the further development of PRS.

# Multi-tissue differential expression analyses between proxy case and control samples

By

Yutong Liu

B.S.A., Zhejiang University, 2021

Thesis Committee Chair: Steve Qin, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of

Master of Science in Public Health in Biostatistics
2023

## Table of Contents

## 1. Introduction

When dealing with a polygenic disorder, a combination of multiple risk variants is necessary to obtain an adequate measure that can effectively identify individuals at a high risk of developing the disease. [1] There are numerous methods available for consolidating information from various gene locations, while the most used one is the polygenic risk score (PRS). PRS is a score uses a weighted sum of allele dosages multiplied by their corresponding effect sizes to assess the risk of having or developing a particular medical condition in each individual. In addition to identifying shared causes of traits, PRS have been employed to test for genome-wide interactions between genes and environmental factors or between different genes [2], to conduct Mendelian Randomization studies to infer causal relationships, and also useful for patient stratification and sub-phenotyping [3]. A PRS typically could contain hundreds-to-millions of genetic variants, which makes it usually achieve substantially greater predictive power than a small number of genome-wide significant Single Nucleotide Polymorphisms (SNPs). [1] There have been 3361 PRS established in the Polygenic Score (PGS) Catalog, and there is a shift from utilizing them solely in research discovery studies towards clinical research studies. Evidence supporting the transition comes widely from cardiovascular disease, type 2 diabetes, breast and prostate cancers, as well as Alzheimer's disease (AD). However, there is still a long route to be covered before PRS can be considered practical instruments for clinicians. [4]

Alzheimer's disease (AD) is one of the most common forms of neurodegenerative disorder, which is estimated affecting 47 million people around the world. [5] Family history is studied to be the second great risk factor after aging for AD, which is categorized into two forms: the

early-onset familial cases and the late-onsite cases. [6] Although the familial clustering is found more obvious in the former one, the heritability for late-onsite Alzheimer's disease is estimated up to 58%-79%. [7] Amyloid precursor protein gene (*APP*) and presenilin genes (*PSEN1*, *PSEN2*) are found related to early-onset familial cases, while the apolipoprotein E gene (*APOE*) is found associated with the risk of late-onsite Alzheimer's disease. [8] With the development of the genome-wide association study (GWAS), *CLU*, *PICALM* and *CR1* are found showing genome-wide statistically significant association with AD. [8, 9] Now, there are over 50 risk loci with genome-wide significance (GWS; $P < 5 \times 10^{-8}$) found associated with AD. [9] Thanks to high-throughput genomic approaches, our understanding of the genetics of AD has advanced considerably in the past decade. However, there is still much work to be done to identify the absent genetic elements and to differentiate the credible findings from the less reliable ones. Nowadays, more and more computational methods are applied to unmask novel genes related to AD based on information we got from GWAS, such as machine learning with random forest and LASSO [10], and artificial intelligence-driven meta-analysis [11].

As for this study, I'd like to conduct differential expression analysis across human brain tissues based on PRS to learn about how different people with higher and lower score are in their expressions and how well PRS works in defining proxy case and control samples.

## 2. Methods

## 2.1 **Data Source**

The present study employed data obtained from the Genotype-Tissue Expression (GTEx) database, a comprehensive resource that integrates tissue banking and genetic analysis to facilitate investigations into the relationship between genetic variation and gene expression as well as other molecular phenotypes across numerous reference tissues.[12] The database encompasses information derived from 54 non-diseased tissue sites, incorporating data from approximately 900 individuals. Specifically, this study utilized genotype and expression data extracted from whole blood and 13 distinct brain tissue.

## 2.2 **PRS calculation**

This study employs PRS comprising 19 SNPs that have been strongly associated with AD in previous research.[13] PRS scores for each participant are generated by weighting their genotype at each SNP by the corresponding log odds ratios provided in the PRS. (Table 1)

## 2.3 **Differential Expression Analysis**

The study participants were categorized into different groups based on the calculated PRS. The individuals with the highest and lowest 10% PRS scores in each tissue were selected for differential expression analysis. The analysis was performed using the DESeq2 package in R (4.0.0). This analysis was conducted independently for all fourteen tissues, and the results were illustrated using volcano plots for each tissue.

We identified genes that exhibited a significant increase or decrease in expression levels by at least a two-fold magnitude from all the thirteen brain tissues (adjusted p-value $< 0.05$). Subsequently, we utilize a heatmap to assess their significance in each tissue. To exclude the

influence of outliers, we generated some boxplots between the PRS and their expression counts for the genes with the highest significance. Moreover, we conducted further investigations into the genes that displayed notable up or down-regulation in particular tissues, exploring their relationship with AD.

### 2.4 **Database Search and Chi-Square Test**

DisGeNET is a discovery platform that houses one of the most extensive collections of publicly available genes and variants linked to human diseases. [14] This platform includes 1,134,942 gene-disease associations (GDAs) between 21,671 genes and 30,170 diseases, disorders, traits, clinical or abnormal human phenotypes. Among the genes present in the database, 3397 are associated with Alzheimer's disease.

In the present study, we utilized the DisGeNET database about AD to conduct a chi-square analysis to evaluate the extent of enrichment of AD-related genes that were identified through differential expression analysis based on the PRS in each tissue. Before testing, all genes were transformed into symbol format based on the GENCODE database, and any genes not annotated by GENCODE were eliminated. The chi-square test was utilized to assess whether the proportion of AD-related genes was higher in the top 1000 most significant genes from the differential expression analysis, as compared to all genes. The analysis was performed using R software version 4.0.0, with a significance threshold set at 0.05.

## 3. Results

### 3.1 **Distribution of PRS Score**

866 samples from GTEx are included in the score calculation with 3 samples without one of the SNPs information. The score is roughly normally distributed, with mean of 2.63 and standard deviation 0.33. (Figure 1)

In all the brain tissues, samples available for differential expression analysis are limited, with the least from brain amygdala of 26, and the most from brain cerebellum, brain cortex, and brain nucleus accumbens (basal ganglia) of 42. (Figure 2) While in whole blood, 134 samples in total are recruited into the differential expression analysis, with half in higher PRS group and half in lower PRS group. (Figure 2)

### 3.2 **Differential Expression Analysis in Each Tissue**

In the differential expression analysis across all the tissues, only a few genes show a significant up or down-regulation in expression level by at least a two-fold magnitude. (Figure 3) In the brain tissues, we found the most of significant genes in the spinal cord (cervical c-1) (Figure 4.f), while the nucleus accumbens (basal ganglia) has no significant genes (Figure 4.j), which is the least.

Few genes are found significant across all the brain tissues, and only 7 genes are found significant in 2 of the 13 tissues, which are *CXCL10*, *HSPA6*, *NPAS4*, *MTCO1P12*, *SORD2P*, *HIST1H2BG*, *HIST1H2AE*. (Figure 5) There are 154 genes found significant from all the brain tissues, and among all these significant genes, eight mitochondrial genes and two histone encoding genes. (Figure 5)

In whole blood, there are eighteen genes that exhibited a significant increase or decrease in expression levels by at least a two-fold magnitude, in which three (*IL22RA2*, *LGR4*, *SMPDL3A*) are up-regulated and the other fifteen are down-regulated. In all the eighteen significant genes, *MTCO1P12* is also detected significant in brain anterior cingulate cortex (BA24) and brain caudate basal ganglia; *SORD2P* is also detected significant in brain hippocampus and brain cortex.

In the boxplots of the fifteen most significant genes, some genes are showing significant because of some outliers, such as *OASL* (Figure 6.c), *CXCL10* (Figure 6.d), *PITX2* (Figure 6.e), *HSPA6* (Figure 6.g), *NPAS4* (Figure 6.h), while others are showing expression differences in the two groups with different levels of PRS. In these genes, some of them are found related to AD in some way.

For *MT-TL2* (Figure 6.a), it has been found some association between mitochondrial DNA variations and AD [15], and it's also been found in Harmonizome [16] database related to AD which has 1462 related genes in total. *POMC* (Figure 6.b)controls the synthesis of pro-opiomelanocortin (POMC) whose activation could rescue the impairment in hippocampus-dependent synaptic plasticity in the APP/PS1 mouse model of AD. [17] *HTR2C* (Figure 6.m) is found to be related to the depression of AD [18], *CARTPT* (Figure 6.n) is found related to both obesity and AD [19]. In recent years, with more machine learning and GWAS techniques, *KLHL7-AS1* (Figure 6.f) [20], *COL24A1* (Figure 6.k) [10], *SOX14* (Figure 6.o) [21] are also found related to AD, although the mechanism involved is not clear. *KLHL7-AS1* is also found highly related to Parkinson's disease (PD) [22], which may be related to the comorbid traits between AD and PD [9].

### 3.3 **Database Search and Chi-Square Test**

In all the 14 tissues used for chi-square analysis, only in two we failed to reject that the proportion of AD-related genes is the same in the top 1000 most significant genes from the differential expression analysis, which are brain amygdala ($p = 0.668$), and brain cortex ($p = 0.267$). (Table 2) In seven of all the tissues, the proportion of AD-related genes is highly significantly higher in the top 1000 most significant genes from the differential expression analysis ($p < 0.001$), which are brain anterior cingulate cortex (BA24), brain caudate (basal ganglia), brain frontal cortex (BA9), brain hypothalamus, brain nucleus accumbens (basal ganglia), brain spinal cord (cervical c-1), and brain substantia nigra. (Table 2)

## 4. Discussion

The Polygenic Risk Score (PRS) is a promising tool for defining proxy case and control samples for Alzheimer's disease (AD) and identifying novel AD-associated genes. Nonetheless, our study demonstrated a lower-than-expected number of significant genes identified through differential expression analysis, potentially attributed to the exclusion of *APOE*-related SNPs in the PRS selected. Thus, stratification by *APOE* genotype may enhance the accuracy of gene discovery using PRS. Furthermore, the limited sample size of brain tissues may have hindered the identification of well-established AD-related genes, such as *APOE*, *CLU*, *PICALM*, and *CR1*.

Although our investigation of 18 significant genes discovered in whole blood did not reveal any potential biomarkers for AD prediction or diagnosis, our analysis of 13 brain tissues disclosed several genes with substantially different expression levels in individuals with

higher versus lower PRS scores. These findings include some previously identified AD-related genes with clear mechanisms, such as mitochondrial DNA, which may be linked to the shifting of cell dynamics and facilitating neuronal vulnerability, and *POMC*, which controls the synthesis of pro-opiomelanocortin that could alleviate the impairment in hippocampus-dependent synaptic plasticity. Furthermore, several genes are linked to AD through machine learning and GWAS techniques lack clear mechanisms, while others are novel or even unannotated. They are potentially associated with unexplored AD mechanisms, which worth further investigations.

Although PRS is effective in enriching for AD-related genes in the overall dataset across multiple tissues, there are still unrelated genes, even accounting for mechanisms that have yet to be revealed.

Overall, our findings suggest that PRS is a valuable approach for defining proxy case and control samples for AD. However, the development of PRS with an expanded number of SNPs may improve its predictive power and reduce the inclusion of unrelated genes with high significance. Additionally, we aim to broaden our research pipeline to include other diseases, such as Parkinson's Disease, cardiovascular disease, and diabetes, to further explore the potential of PRS in multi-tissue differential expression analysis and disease-related gene identification.

## Reference


1. Choi, S.W., T.S. Mak, and P.F. O'Reilly, *Tutorial: a guide to performing polygenic risk score analyses.* Nat Protoc, 2020. **15**(9): p. 2759-2772.

2. Agerbo, E., et al., *Polygenic Risk Score, Parental Socioeconomic Status, Family History of Psychiatric Disorders, and the Risk for Schizophrenia: A Danish Population-Based Study and Meta-analysis.* JAMA Psychiatry, 2015. **72**(7): p. 635-41.

3. Mavaddat, N., et al., *Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes.* Am J Hum Genet, 2019. **104**(1): p. 21-34.

4. Lewis, C.M. and E. Vassos, *Polygenic risk scores: from research tools to clinical instruments.* Genome Medicine, 2020. **12**(1): p. 44.

5. Leonenko, G., et al., *Polygenic risk and hazard scores for Alzheimer's disease prediction.* Ann Clin Transl Neurol, 2019. **6**(3): p. 456-465.

6. Tanzi, R.E., *The genetics of Alzheimer disease.* Cold Spring Harb Perspect Med, 2012. **2**(10).

7. Gatz, M., et al., *Role of Genes and Environments for Explaining Alzheimer Disease.* Archives of General Psychiatry, 2006. **63**(2): p. 168-174.

8. Bellenguez, C., B. Grenier-Boley, and J.-C. Lambert, *Genetics of Alzheimer's disease: where we are, and where we are going.* Current Opinion in Neurobiology, 2020. **61**: p. 40-48.

9. Sims, R., M. Hill, and J. Williams, *The multiplex model of the genetics of Alzheimer's disease.* Nature Neuroscience, 2020. **23**(3): p. 311-322.

10. Sharma, A. and P. Dey, *A machine learning approach to unmask novel gene signatures and prediction of Alzheimer's disease within different brain regions.* Genomics, 2021. **113**(4): p. 1778-1789.

11. Finney, C.A., et al., *Artificial intelligence-driven meta-analysis of brain gene expression identifies novel gene candidates and a role for mitochondria in Alzheimer's disease.* Computational and Structural Biotechnology Journal, 2023. **21**: p. 388-400.

12. Lonsdale, J., et al., *The Genotype-Tissue Expression (GTEx) project.* Nature Genetics, 2013. **45**(6): p. 580-585.

13. Chouraki, V., et al., *Evaluation of a Genetic Risk Score to Improve Risk Prediction for Alzheimer's Disease.* J Alzheimers Dis, 2016. **53**(3): p. 921-32.

14. Piñero, J., et al., *The DisGeNET cytoscape app: Exploring and visualizing disease genomics data.* Computational and Structural Biotechnology Journal, 2021. **19**: p. 2960-2967.

15. Lakatos, A., et al., *Association between mitochondrial DNA variations and Alzheimer's disease in the ADNI cohort.* Neurobiol Aging, 2010. **31**(8): p. 1355-63.

16. Rouillard, A.D., et al., *The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins.* Database, 2016. **2016**: p. baw100.

17. Lau, J.K.Y., et al., *Melanocortin receptor activation alleviates amyloid pathology and glial reactivity in an Alzheimer's disease transgenic mouse model.* Sci Rep, 2021. **11**(1): p. 4359.

18. Cacabelos, R., et al., *Pharmacogenetics of anxiety and depression in Alzheimer's disease.* Pharmacogenomics, 2023. **24**(1): p. 27-57.

19. Zhuang, Q.S., et al., *Associations Between Obesity and Alzheimer's Disease: Multiple Bioinformatic Analyses.* J Alzheimers Dis, 2021. **80**(1): p. 271-281.

20. Monk, B., et al. *A Machine Learning Method to Identify Genetic Variants Potentially Associated With Alzheimer's Disease*. Frontiers in genetics, 2021. **12**, 647436 DOI: 10.3389/fgene.2021.647436.

21. Jun, G., et al., *A novel Alzheimer disease locus located near the gene encoding tau protein.* Mol Psychiatry, 2016. **21**(1): p. 108-17.

22. Murthy, M.N., et al., *Increased brain expression of GPNMB is associated with genome wide significant risk for Parkinson's disease on chromosome 7p15.3.* Neurogenetics, 2017. **18**(3): p. 121-133.

Table 1. Polygenic risk score (PRS) used for analysis.

| Locus name | rsID | Chr name | Effect allele | Other allele | Effect weight | Allele frequency effect |
|---|---|---|---|---|---|---|
| CR1 | rs6656401 | 1 | A | G | 0.166 | 0.2 |
| BIN1 | rs6733839 | 2 | T | C | 0.199 | 0.41 |
| INPP5D | rs35349669 | 2 | T | C | 0.077 | 0.49 |
| MEF2C | rs190982 | 5 | A | G | 0.077 | 0.59 |
| HLA-DRB5/HLA-DRB1 | rs9271192 | 6 | C | A | 0.104 | 0.28 |
| CD2AP | rs10948363 | 6 | G | A | 0.095 | 0.27 |
| NME8 | rs2718058 | 7 | A | G | 0.077 | 0.63 |
| ZCWPW1 | rs1476679 | 7 | T | C | 0.095 | 0.71 |
| EPHA1 | rs11771145 | 7 | G | A | 0.104 | 0.66 |
| PTK2B | rs28834970 | 8 | C | T | 0.095 | 0.37 |
| CLU | rs9331896 | 8 | T | C | 0.148 | 0.62 |
| CELF1 | rs10838725 | 11 | C | T | 0.077 | 0.32 |
| MS4A6A | rs983392 | 11 | A | G | 0.104 | 0.6 |
| PICALM | rs10792832 | 11 | G | A | 0.140 | 0.64 |
| SORL1 | rs11218343 | 11 | T | C | 0.262 | 0.96 |
| FERMT2 | rs17125944 | 14 | C | T | 0.131 | 0.09 |
| SLC24A4/RIN3 | rs10498633 | 14 | G | T | 0.095 | 0.78 |
| ABCA7 | rs4147929 | 19 | A | G | 0.140 | 0.19 |
| CASS4 | rs7274581 | 20 | T | C | 0.131 | 0.92 |

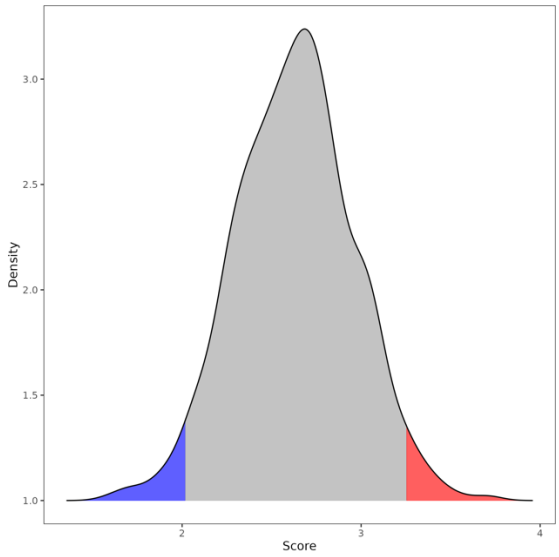



Figure 1. The distribution of polygenic risk score (PRS) of 863 samples from GTEx.

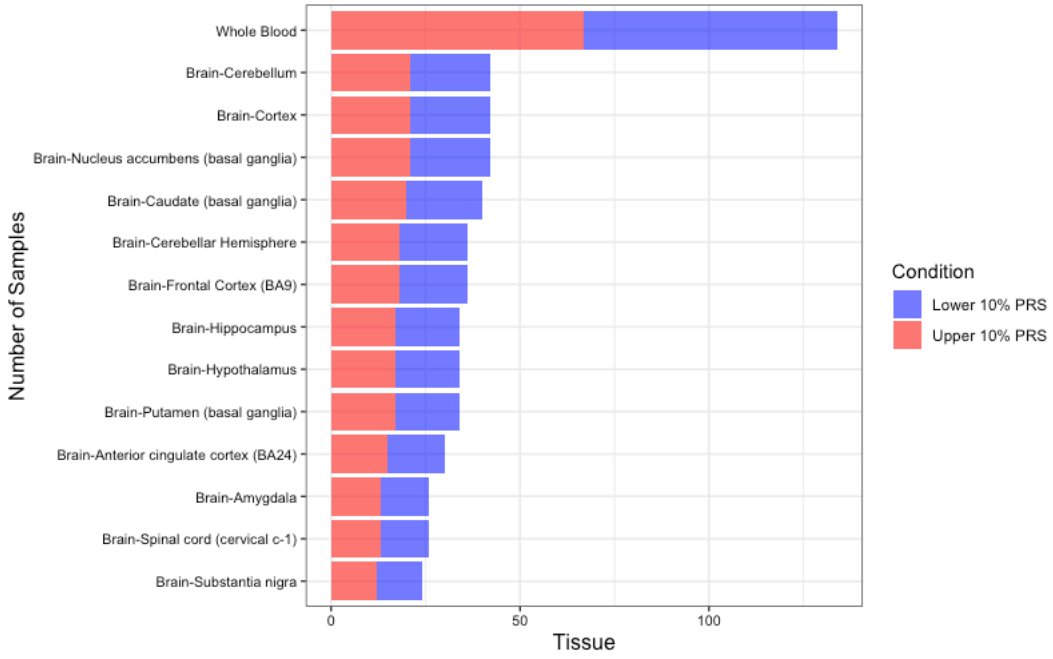



Figure 2. Number of Samples in Each Tissue Used for Differential Expression Analysis.

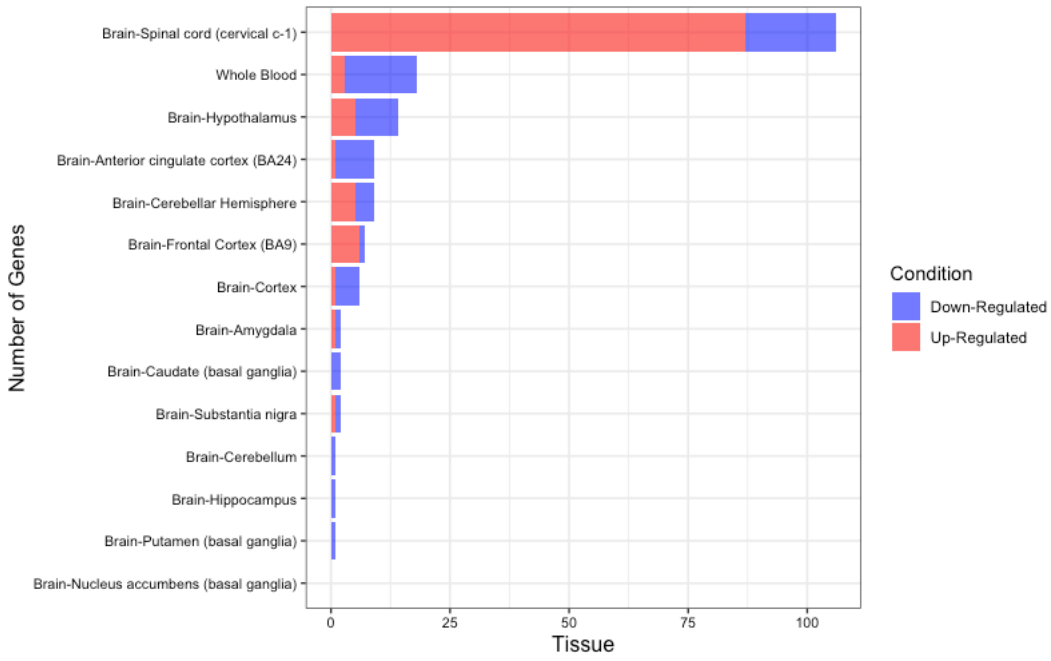
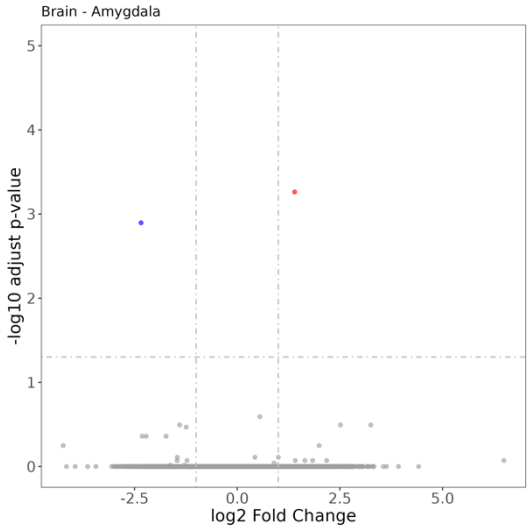
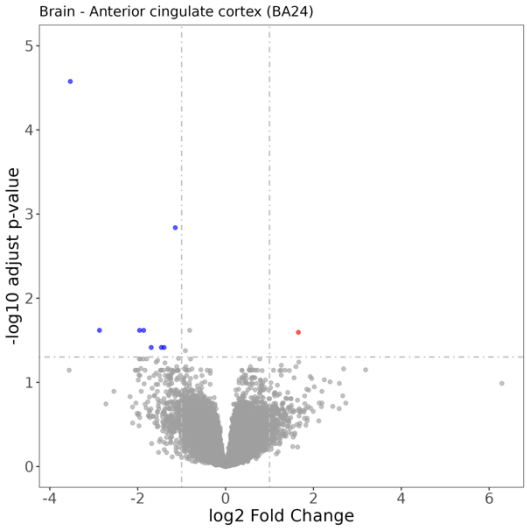


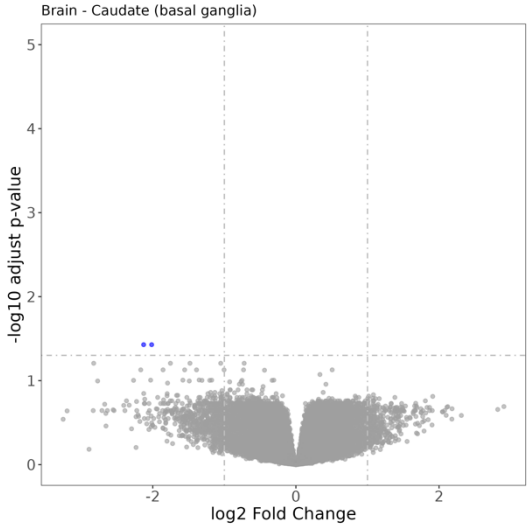

Figure 3. Number of Gene Significantly Up or Down Regulated by at least Two Folds.


Brain - Amygdala
-log10 adjust p-value
log2 Fold Change

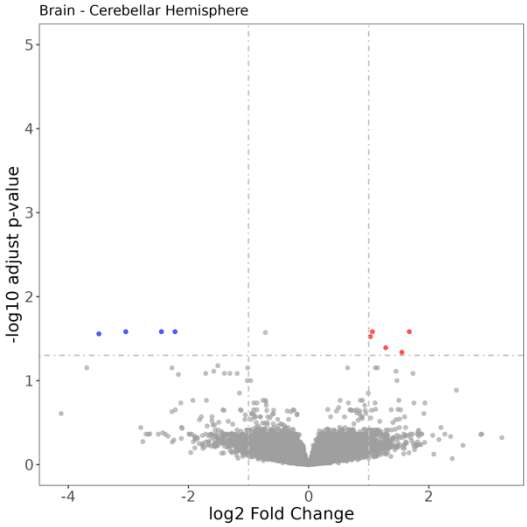
a


Brain - Anterior cingulate cortex (BA24)
-log10 adjust p-value
log2 Fold Change

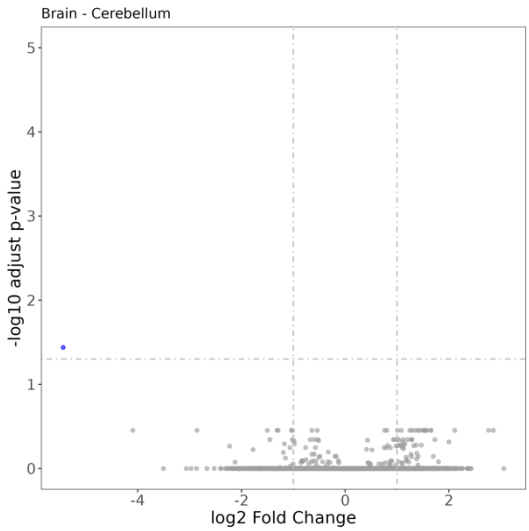
b

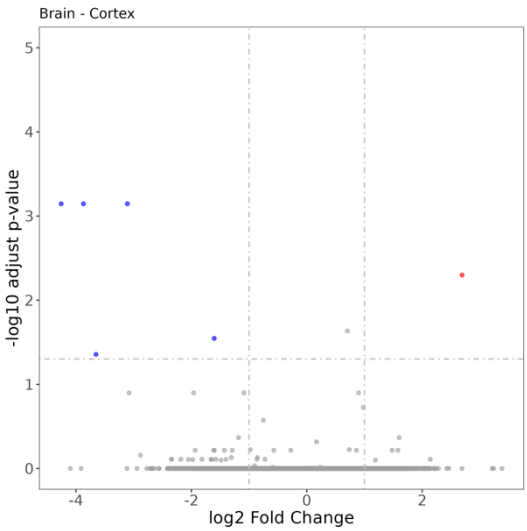

Brain - Caudate (basal ganglia)
-log10 adjust p-value
log2 Fold Change

c


Brain - Cerebellar Hemisphere
-log10 adjust p-value
log2 Fold Change

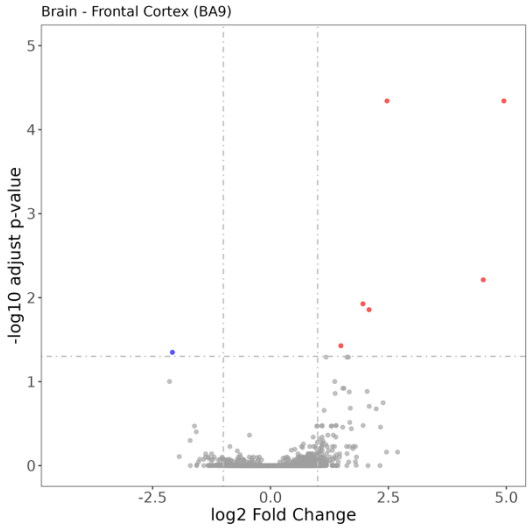
d

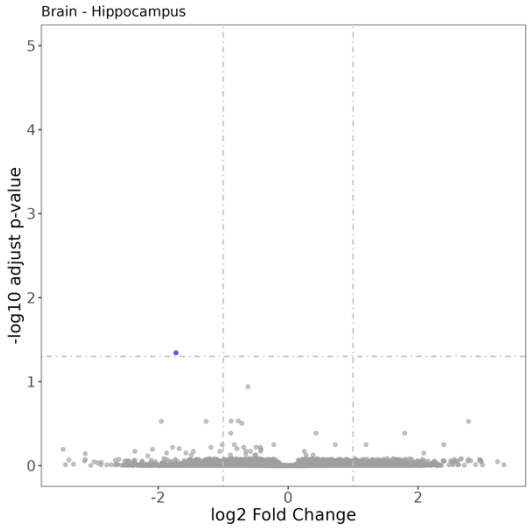

Brain - Cerebellum
-log10 adjust p-value
log2 Fold Change


Brain - Cortex
-log10 adjust p-value
log2 Fold Change

e

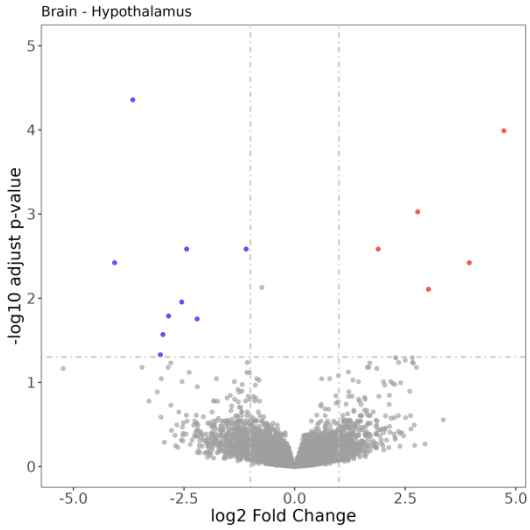
f

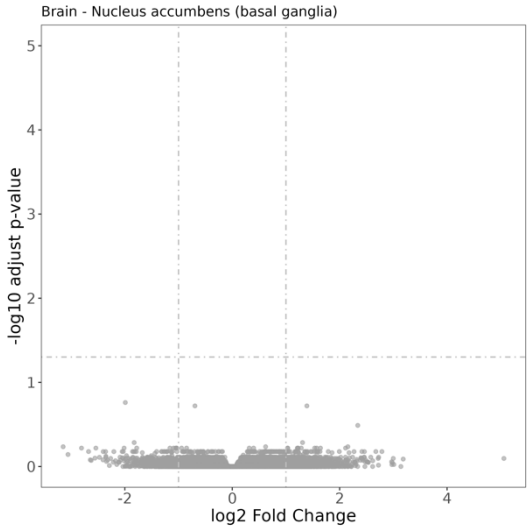

Brain - Frontal Cortex (BA9)
-log10 adjust p-value
log2 Fold Change
0
1
2
3
4
5
-2.5
0.0
2.5
5.0


Brain - Hippocampus
-log10 adjust p-value
log2 Fold Change
0
1
2
3
4
5
-2
0
2

g

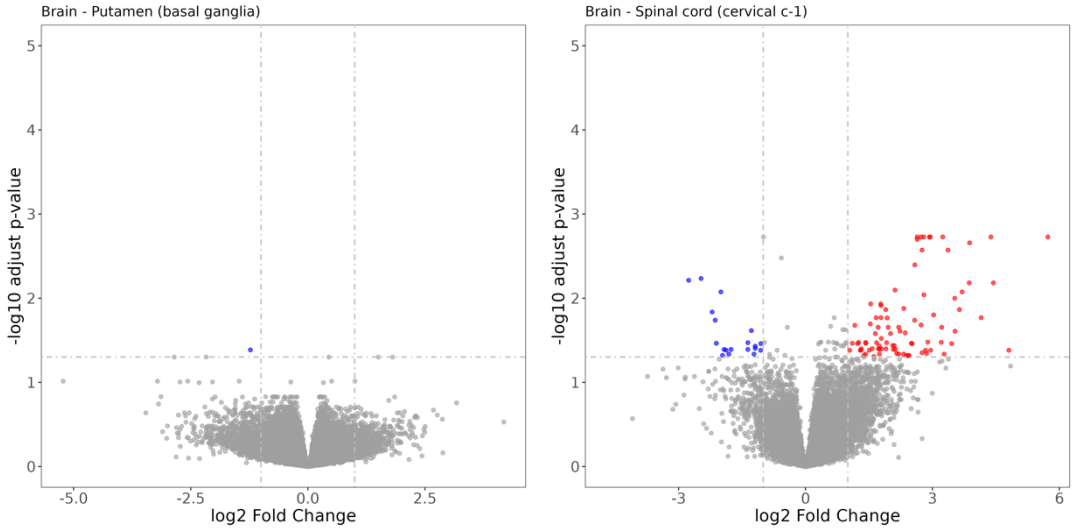
h


Brain - Hypothalamus
-log10 adjust p-value
log2 Fold Change
0
1
2
3
4
5
-5.0
-2.5
0.0
2.5
5.0

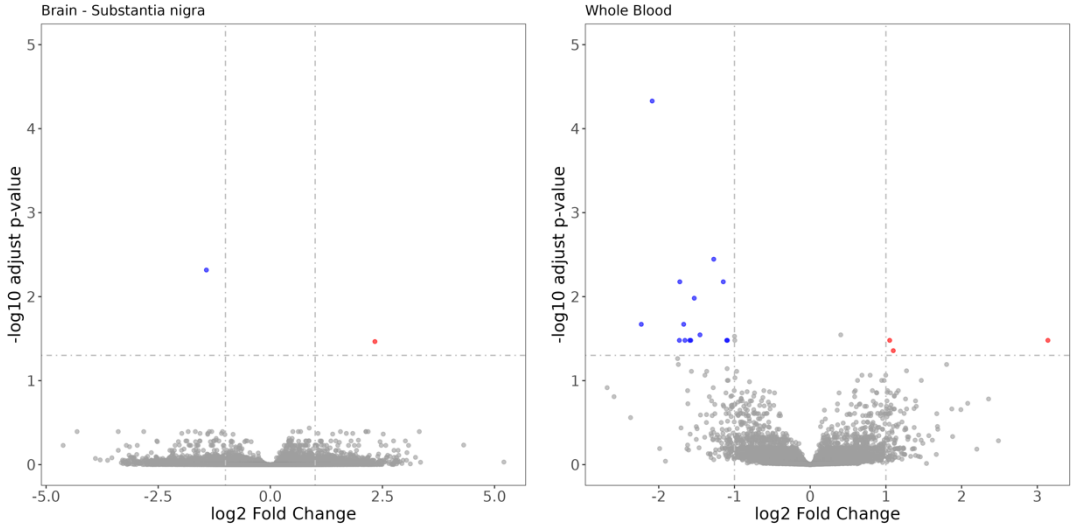

Brain - Nucleus accumbens (basal ganglia)
-log10 adjust p-value
log2 Fold Change
0
1
2
3
4
5
-2
0
2
4

i

j




k l




m n

Figure 4. Volcano Plots for Differential Expression in Each Tissue.

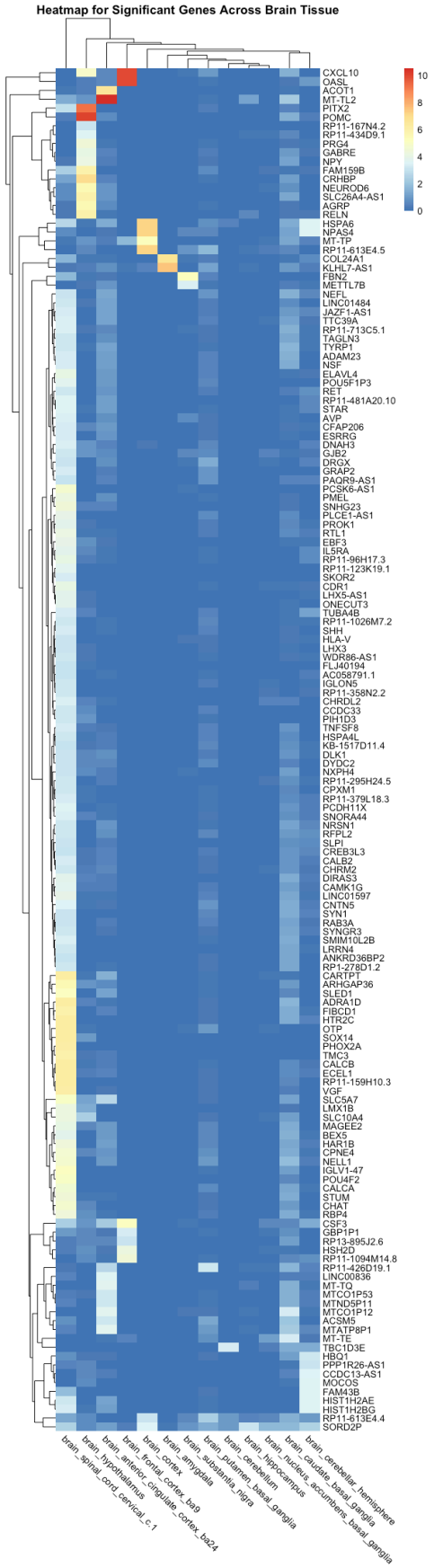
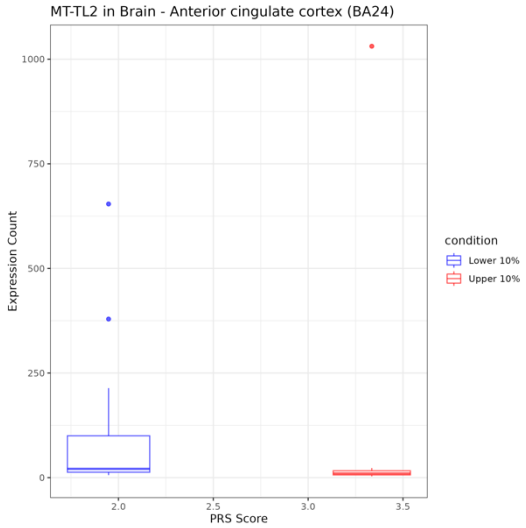
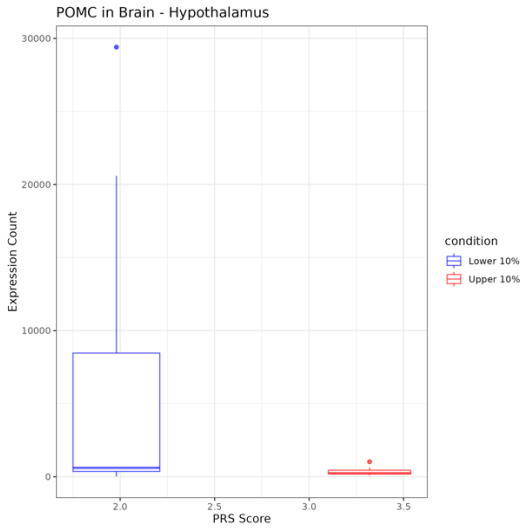


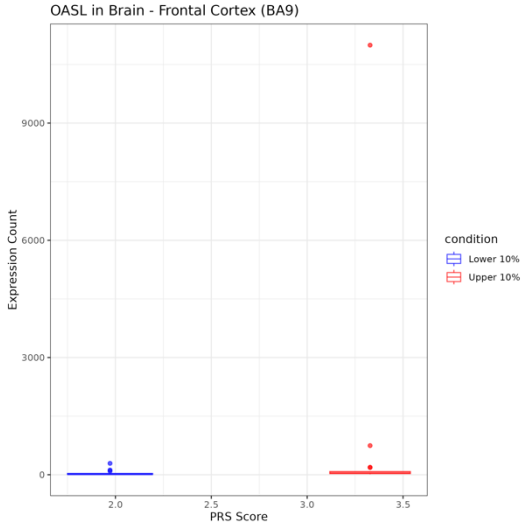

Figure 5. Heatmap for the significance of significant genes across all the brain tissues.


MT-TL2 in Brain - Anterior cingulate cortex (BA24)
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

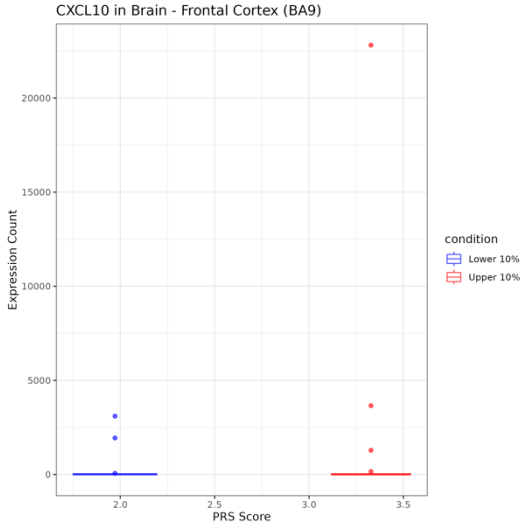
a


POMC in Brain - Hypothalamus
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

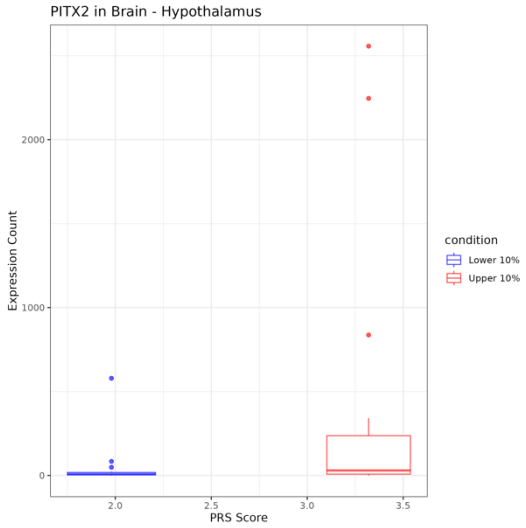
b


OASL in Brain - Frontal Cortex (BA9)
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

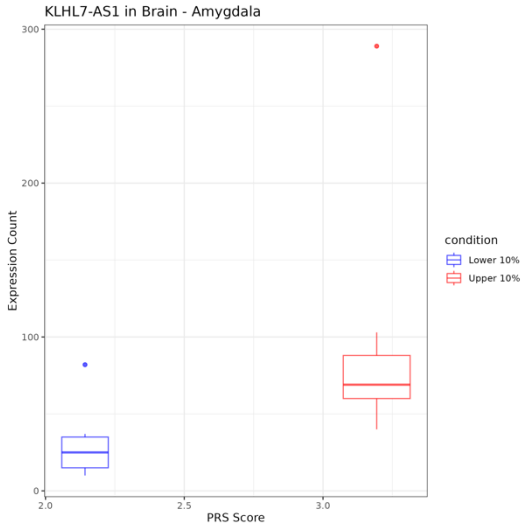
c


CXCL10 in Brain - Frontal Cortex (BA9)
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

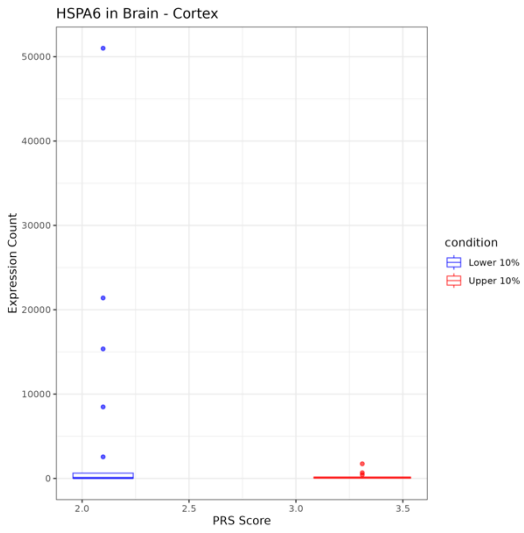
d


PITX2 in Brain - Hypothalamus
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

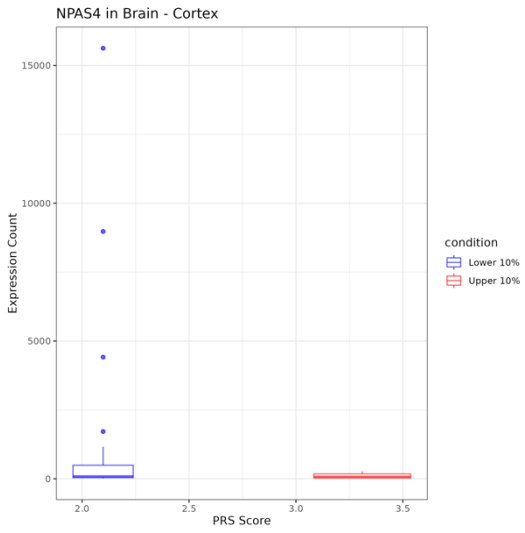
e


KLHL7-AS1 in Brain - Amygdala
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

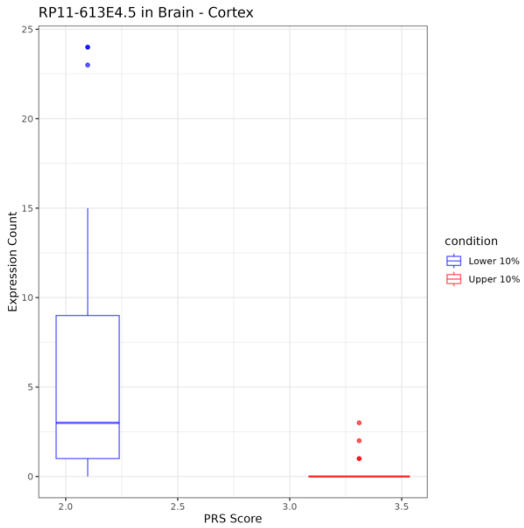
f


HSPA6 in Brain - Cortex
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

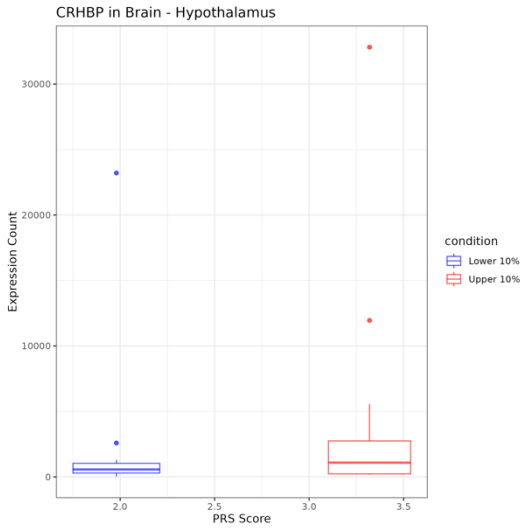
g


NPAS4 in Brain - Cortex
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

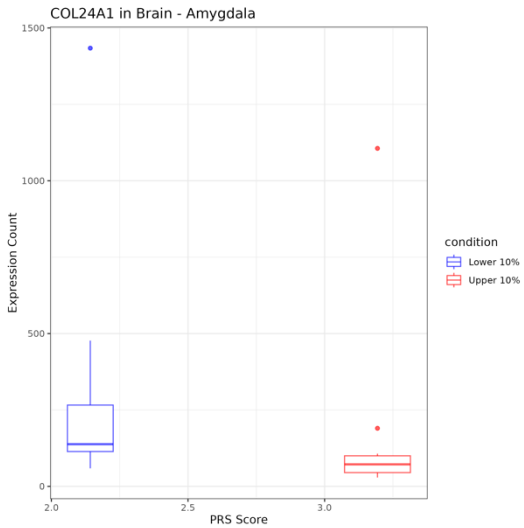
h


RP11-613E4.5 in Brain - Cortex
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

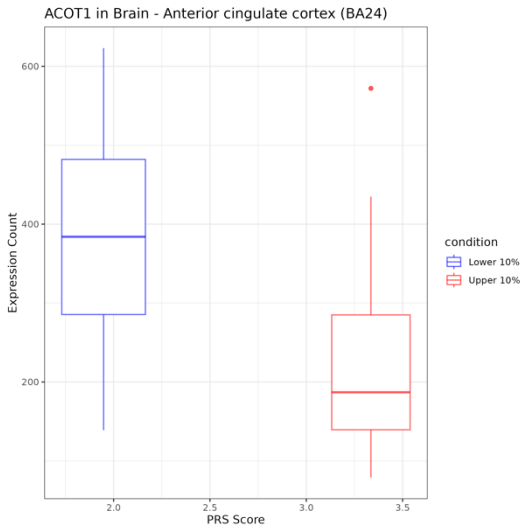
i


CRHBP in Brain - Hypothalamus
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

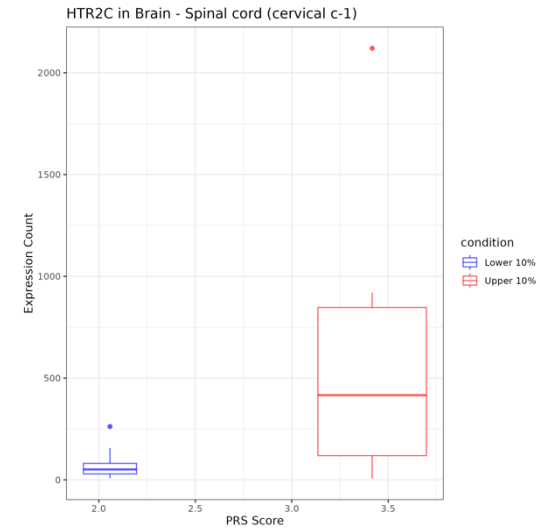
j


COL24A1 in Brain - Amygdala
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

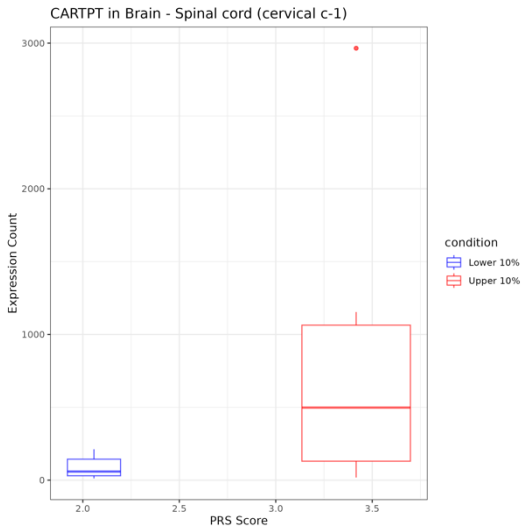
k


ACOT1 in Brain - Anterior cingulate cortex (BA24)
Expression Count
PRS Score
condition
Lower 10%
Upper 10%

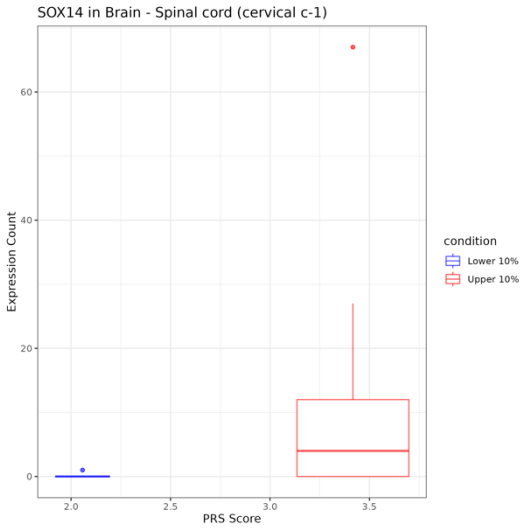
l




m




n




o

Figure 6. Boxplots between PRS and Expression counts in 15 most significant genes.

Table 2. Results for chi-square tests in all the tissues.

| Tissue | Top 1000 DE genes related to AD | Top 1000 DE genes unrelated to AD | All the other DE genes related to AD | All the other DE genes unrelated to AD | Chi-Square | P-value |
|---|---|---|---|---|---|---|
| Brain - Spinal cord (cervical c-1) | 172 | 816 | 2913 | 30259 | 85.88 | <0.001 |
| Brain - Anterior cingulate cortex (BA24) | 148 | 845 | 2952 | 31501 | 47.76 | <0.001 |
| Brain - Frontal Cortex (BA9) | 144 | 844 | 2946 | 31944 | 45.11 | <0.001 |
| Brain - Caudate (basal ganglia) | 136 | 854 | 2966 | 32430 | 34.77 | <0.001 |
| Brain - Substantia nigra | 141 | 850 | 2927 | 28745 | 27.49 | <0.001 |
| Brain - Hypothalamus | 131 | 863 | 2975 | 31910 | 25.85 | <0.001 |
| Brain - Nucleus accumbens (basal ganglia) | 119 | 870 | 3001 | 33113 | 16.84 | <0.001 |
| Brain - Putamen (basal ganglia) | 117 | 872 | 2962 | 30478 | 10.06 | 0.002 |
| Brain - Hippocampus | 112 | 875 | 3006 | 31748 | 8.44 | 0.004 |
| Whole Blood | 105 | 888 | 3023 | 33762 | 6.76 | 0.009 |
| Brain - Cerebellar Hemisphere | 107 | 880 | 2992 | 32148 | 6.33 | 0.012 |
| Brain - Cerebellum | 101 | 887 | 3001 | 32805 | 3.99 | 0.046 |
| Brain - Cortex | 93 | 895 | 3016 | 33022 | 1.23 | 0.267 |
| Brain - Amygdala | 92 | 896 | 3002 | 30856 | 0.18 | 0.668 |