UNIVERSITY OF CALIFORNIA SAN DIEGO

Systems Level Analysis of Late Onset Alzheimer's Disease Across Four Brain Regions

A thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Bioengineering

by

Katherine Ann Pieri

Committee in charge:

Professor Shankar Subramaniam, Chair
Professor Ratnesh Lal
Professor Steven Wagner

2020

The thesis of Katherine Ann Pieri is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California San Diego

2020

DEDICATION

This thesis is dedicated to my grandmothers.

Jean Ann Carroll

&

Annabelle Lois Pieri

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'd like to express my gratitude to Dr. Shankar Subramaniam. Thank you for your guidance and encouragement of my thesis and education. What I have learned in your lab has been invaluable.

I'd also like to express my appreciation for Dr. Andrew Caldwell. Thank you for your endless support and advice; it has made all the difference in the success of this thesis. Your guidance and thoughtful feedback have helped to shape my learning over the past two years.

I am also very grateful for my friends Ellie and Priya, for their never-ending encouragement. I would also like to acknowledge my family for all of their support over the years. Lastly, I'd like to give a special thanks to Colin for his endless encouragement over the past decade; his support has been immeasurable.

ABSTRACT OF THE THESIS

Systems Level Analysis of Late Onset Alzheimer's Disease Across Four Brain Regions

by

Katherine Ann Pieri

Master of Science in Bioengineering

University of California San Diego, 2020

Professor Shankar Subramaniam, Chair

Alzheimer's Disease (AD) is a neurodegenerative disease that causes loss of memory, among other cognitive functions and is the 6th leading cause of death in the United States. Late Onset Sporadic Alzheimer's Disease (LOAD) is the most common form of AD and yet still little is known about the disease and there is no cure. The disease is complex, with temporally distinct effects at different stages. Microarray data generated from 885 postmortem patient brain tissue samples (401 control samples, 484 AD samples) was used to study the effects of LOAD on four

regions of the brain: the dorsolateral Prefrontal Cortex (PFC), Visual Cortex (VC), Cerebellum (CR) and the Middle Temporal Gyrus (MTG). Various methods were explored and selected to create an analysis pipeline. The use of gene set and transcription factor enrichment enabled functional class sorting that suggested key biological functions affected by AD in each region. Five main endotypes were identified and further analyzed: Cell Cycle (G_1-S Phase), Inflammation, Dedifferentiation, Synaptic Signaling and Mitochondrial Metabolism. Custom gene sets were used to create protein-protein interaction networks in order to identify key TF and gene activity within each brain region, revealing that these five endotypes were preferentially enriched in the PFC and VC brain regions.

INTRODUCTION

Alzheimer's Disease (AD) is a neurodegenerative disease that has been estimated to affect more than 5.5 million Americans (Piras, et al., 2019). AD pathology is characterized by accumulation of β-amyloid protein and tau proteins to form plaques and neurofibrillary tangles, respectively, in the brain. Physiological symptoms including memory loss, personality changes, and loss of other cognitive functioning. The most common type of AD is sporadic, Late Onset AD (LOAD), which is defined by the age of onset occurring on or after the age of 65 years old (Wang, Oelze, & Schumacher, 2008). Little is understood about the mechanisms of AD and a cure does not yet exist. Because the disease is complex, affecting different areas of the brain over the course of disease progression, a Systems Biology approach to studying LOAD may offer new insight into the mechanisms of the disease.

In this work, two datasets were used to characterize gene expression in four brain regions: the dorsolateral Prefrontal Cortex (PFC), processes higher order functions, Visual Cortex (VC), processes visual information captured by retinas, Cerebellum (CR), controls balance and motor movement, and Middle Temporal Gyrus (MTG), processes language and facial recognition. Samples were prepared from postmortem brains of AD patients (n = 484) and ND Controls (n = 401). A pipeline was created to perform preprocessing, differential expression, enrichment analysis. The pipeline results were then used to generate protein-protein interaction (PPI) networks (see **Figure 1**).

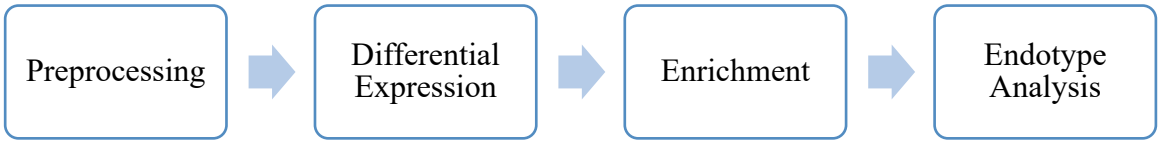


Figure 1. Pipeline Overview

PIPELINE DEVELOPMENT

Data Collection

Two datasets were sourced from Gene Expression Omnibus (GEO). The first dataset was generated from 230 postmortem patient tissues samples from each of three brain regions: PFC, VC, CR (see **Table 1**), using a Rosetta/Merk Human 44k 1.1 microarray platform (data accessible at NCBI GEO database (Zhang, et al., 2013), accession GSE44772). AD subjects were diagnosed at intake and LOAD pathology examinations were performed (Zhang, et al., 2013). The second dataset was generated from 195 postmortem patient tissue samples from the MTG region of the brain (see **Table 2)**, using an Illumina HumanHT-12 V4.0 expression beadchip platform for microarray analysis (data accessible at NCBI GEO database (Piras, et al., 2019), accession GSE132903). Extensive LOAD related pathology examinations were performed on all subjects and LOAD diagnoses were assigned postmortem (Piras, et al., 2019).

Table 1. Post-mortem brain samples taken from 230 patients from each of three brain regions: dorsolateral Prefrontal Cortex (PFC), Visual Cortex (VC) and Cerebellum (CR) (Zhang, et al., 2013).

	NDC	AD	Total
Female	19	67	86
Male	82	62	144
Total	101	129	230

Table 2. Post-mortem brain samples taken from 195 patients from the Middle Temporal Gyrus (MTG) (Piras, et al., 2019).

	NDC	AD	Total
Female	48	48	96
Male	50	49	99
Total	98	97	195

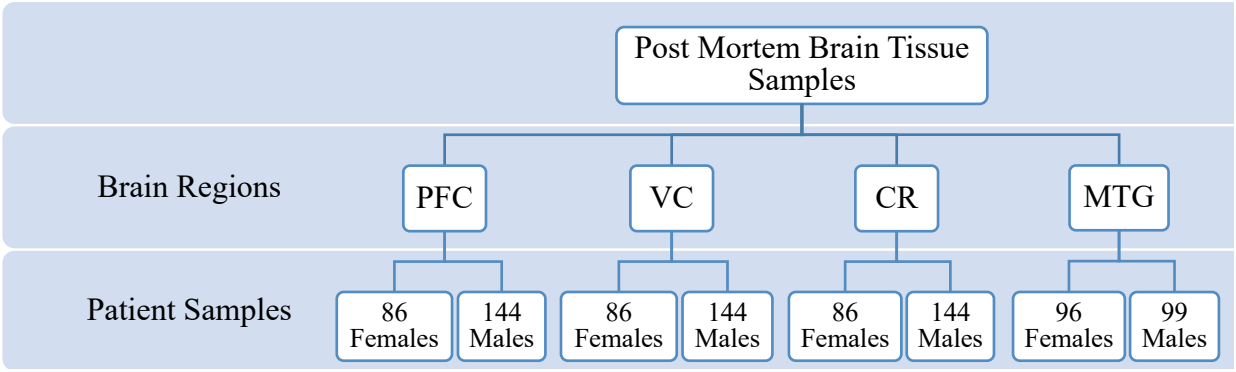


Figure 2. Breakdown of Patient Samples by Brain Region and Sex.

Preprocessing

Normalized microarray data and phenotypic metadata provided on GEO was used for preprocessing: PFC (GSE44770), CR (GSE44768), VC (GSE44771), MTG (GSE132903). The following equation was reportedly used to normalize the PFC, CR and VC data (Zhang, et al., 2013):

$$log_2(\frac{test\ sample}{control\ sample})$$

The MTG test samples were normalized using the lumiExpresso R function, which has a targeted use for transforming raw Illumina probe intensities to expression values (Piras, et al., 2019).

The PFC, CR and VC data was additionally normalized for batch variation. This step was omitted for MTG as information of only one batch was reported. Contrast and design matrices were then created to form three different groups to be analyzed downstream in the pipeline: All Sexes, Females, and Males. Using the *limma* package for R (Smyth, et al., 2020), the data was fit to a linear model before proceeding through the rest of the pipeline.

Differential Expression

Differential expression is a statistical approach used to identifying significant quantitative changes in gene expression levels between experimental groups (i.e., AD vs Control). Many methods utilize p-value cutoffs (e.g., p-value < 0.05) as a measure of significance and effect size, defining such genes as Differentially Expressed Genes (DEGs). While this may sometimes be appropriate, this type of approach can also be limited by the quality of evidence utilized (Harrison, Pattison, Powell, & Beilharz, 2019). For this reason, three different differential expression methods were explored in this pipeline: *eBayes*, *treat* and *topconfects*.

Differential Expression Methods

Empirical Bayes Statistics for Differential Expression - eBayes

This method from *limma* identifies DEGs from a linear model produced by *limma*. *eBayes* reduces sample variances to a common value and calculates moderated t-statistics and moderated F-statistics to identify and rank DEGs. (Ritchie, et al., 2015). The null hypothesis is that all expression values are zero. This corresponds to a log fold change (LFC) of zero, where fold change is the ratio of test to control expression values.

Empirical Bayes Statistics for Differential Expression - treat

This method ranks DEGs using an empirical Bayes method, similar to *eBayes*. However, *treat* uses moderated t-statistic to rank genes against a specified minimum absolute LFC threshold (Ritchie, et al., 2015). The null hypothesis is that all expression values fall within the range of [-LFC, LFC]. As the chosen LFC value approaches zero, the *treat* method converges with the *eBayes* method. *treat* places a greater weight on fold change values than *eBayes* and

allows for more focused insight into the DEGs with greater LFC values. This method was tested using three LFC thresholds: 1.1, 1.2 and 1.5.

TOP results by CONfident efFECT Size (topconfects)

This differential expression method identifies and ranks DEGs using confident effect sizes. This method is a variation of *limma's treat*; however, instead of calculating p-values, *topconfects* calculates a confect score for each gene based on confidence bounds of the LFC given a specified false discovery rate (FDR) (Harrison, Pattison, Powell, & Beilharz, 2019). An FDR of 0.01 was used for all groups.

Differential Expression Results & Method Selection

The *eBayes* and *treat* methods were both initially implemented to explore the distribution of the DEGs. Using *eBayes* resulted in a wider spread of data and a larger number of DEGs, than did using *treat*. The LFC threshold was tested for *treat*, and all LFCs were found to be highly restrictive of the data. At the lowest tested LFC threshold tested, 1.1, relatively few DEGs were identified (see **Table 3**). The results indicated a large number of statistically significant DEGs with small log fold changes. To maintain insight into these genes, the *treat* method was not used moving forward.

Table 3. Counts of DEGs from eBayes and treat (LFC=1.1). (Note: These counts reflect the All-Sexes group. The Female and Male groups both reflected similar results.)

	PFC		MTG		VC		CR	
	eBayes	treat	eBayes	treat	eBayes	treat	eBayes	treat
Up	12283	5740	10883	1732	14380	5961	8936	2334
Not Sig	14180	27864	21945	38737	12238	26769	21216	34168
Down	12817	5676	9351	1710	12662	6550	9128	2778

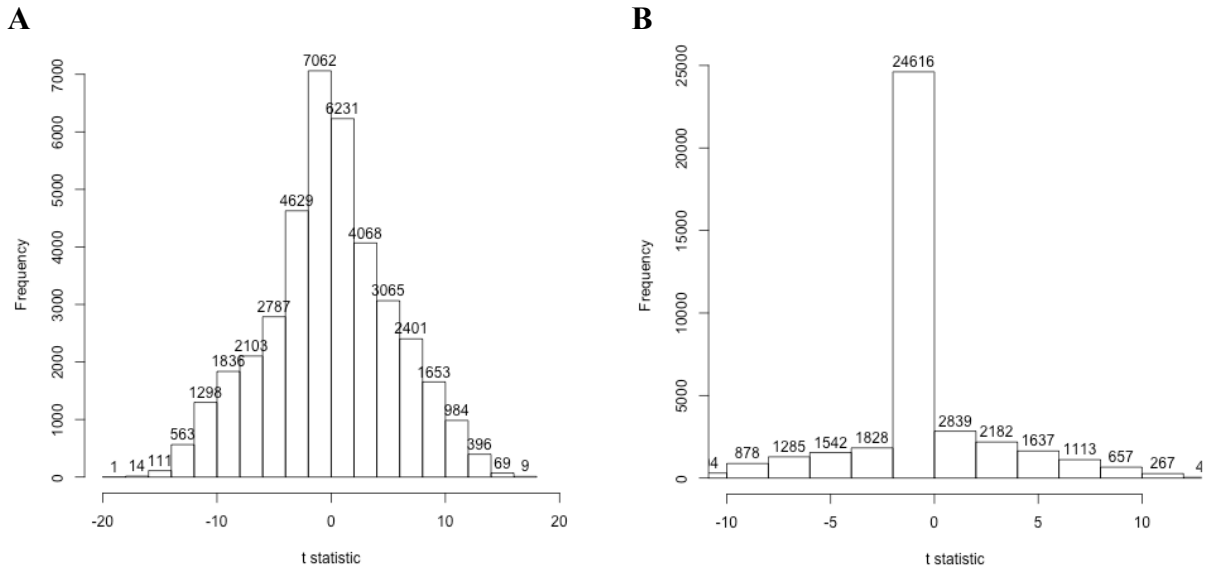


Figure 3. Comparison of eBayes and treat differential expression techniques. A) Histogram of results from *eBayes*, and B) Histogram of the results from *treat*, both applied to the data from the All-Sexes group from the PFC brain region.

The *topconfect*s method resulted in a similar, but ultimately different, ranking of DEGs than generated by *eBayes*. A histogram of confect scores indicated a normal distribution about zero (**Figure 4A**). A confects plot (**Figure 4B**) reveals the top forty genes ranked by confect score. The size and location of the dots represent the average expression value and LFC values as determined by *limma*. The lines indicate the confidence bound for each gene, calculated by *topconfects*. A rank plot was then generated to compare the *topconfects* results to those of *eBayes* (**Figure 4C**). The top forty DEGs as determined by both *topconfects* and *eBayes* are listed with lines connecting genes duplicated by each method. Seventeen of the top forty genes from *eBayes* are within the top genes ranked by *topconfects*, however *topconfects* highly ranked twenty-three different genes. These genes all have a large effect size and sufficient evidence to support this. See **Appendix A** for complete results for All-Sexes, Females and Males in all brain regions.

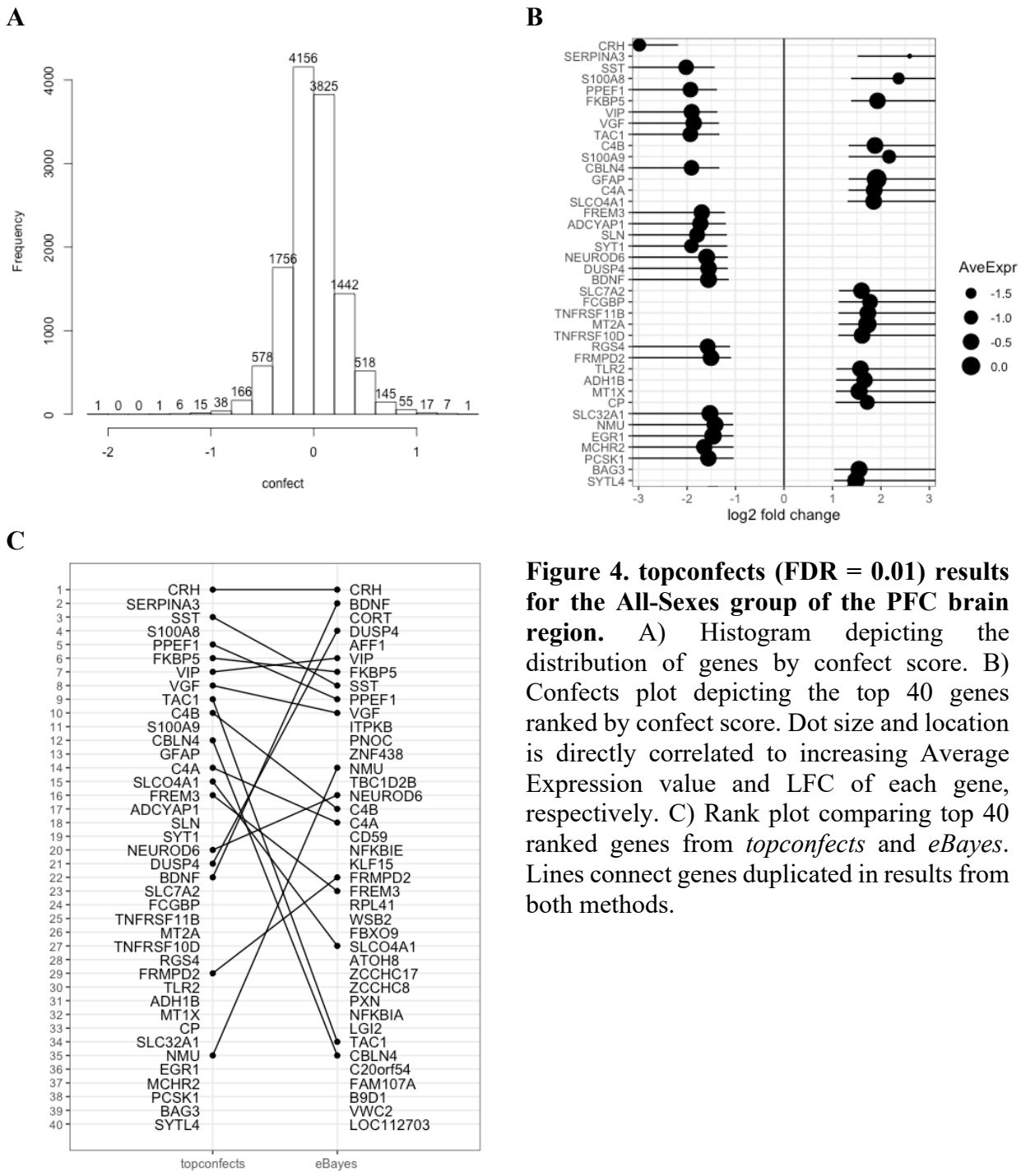


Figure 4. topconfects (FDR = 0.01) results for the All-Sexes group of the PFC brain region. A) Histogram depicting the distribution of genes by confect score. B) Confects plot depicting the top 40 genes ranked by confect score. Dot size and location is directly correlated to increasing Average Expression value and LFC of each gene, respectively. C) Rank plot comparing top 40 ranked genes from *topconfects* and *eBayes*. Lines connect genes duplicated in results from both methods.

Gene Set Enrichment

Gene set enrichment is a common class of methods utilized to gain insightful information from expression of genes at the level of a gene set, where a gene set is a group of genes that

share biological functionality, chromosomal location or regulation. There are a variety of methods used to accomplish this, including Over-Representation Analysis, Functional Class Sorting and Pathway Topology (Zyla, et al., 2019). Within each category of gene set enrichment, there are a variety of tests, with varying approaches. Many methods sort genes by a given metric and may or may not incorporate a gene expression effect size (Zyla, et al., 2019). Given the strengths and weaknesses of the various available enrichment tests, multiple tests were utilized in this pipeline for more comprehensive insight into the gene set functional class of the DEGs.

Gene Set Enrichment Methods

Coincident Extreme Ranks in Numerical Observations (CERNO)

The CERNO enrichment test uses a modified Fisher's combined probability test. It performs enrichment on a list of genes sorted by a given metric. The null hypothesis assumes a random distribution of genes belonging to individual gene sets (Zyla, et al., 2019). One of the features of this test is that it eliminates the need to apply a p-value or LFC threshold. It has also been found to allow for high reproducibility while maintaining speed and precision (Zyla, et al., 2019). Results of this test are functional classes determined by a selected gene set database.

For this test, the *eBayes* object was used as the input. The genes were sorted by Minimal Significant Distance (MSD), defined for positive LFCs to be the lower boundary of the 95% confidence interval, and for negative LFC to be zero minus the upper boundary of the 95% confidence interval (Weiner, 2020). Two gene set databases were used for functional class sorting: Hallmark (Liberzon, Birger, Thorvaldsdottir, Ghandi, & Mesirov, 2015) and Gene Ontology Biological Processes (GOBP) (Ashburner, et al., 2000) (The Gene Ontology Consortium, 2019).

fast Gene Set Enrichment Analysis, Multilevel (fGSEA Multilevel)

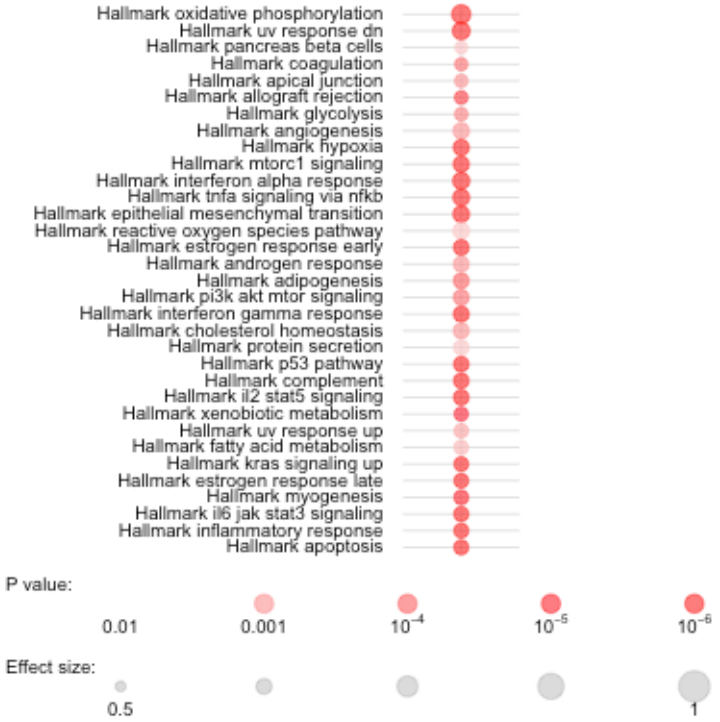
The fGSEA enrichment test is a version of the GSEA enrichment test that utilizes a 'fast' algorithm that enables increased permutations and smaller p-values (Korotkevich, Sukhov, Budin, & Sergushichev, 2020; Korotkevich, Sukhov, & Sergushichev, Fast gene set enrichment analysis, 2019; Subramanian, et al., 2005). The 'multilevel' portion of the algorithm allows for adaptive multilevel splitting Monte Carlo approach. This enrichment test is able to calculate arbitrarily small p-values that allow for improved ranking and subsequent biological insight into the results. For this test, the t-value ranking was used as the input. Two gene set databases were used for functional class sorting: Hallmark and GOBP.

Gene Set Enrichment Results

Coincident Extreme Ranks in Numerical Observations (CERNO)

All CERNO results were ranked by a combination of Effect Size (ES) and p-value. Within the PFC, analyzing the All-Sexes group using CERNO with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Oxidative Phosphorylation (ES = 0.70, p-value = 6.1e-9), Interferon alpha response (ES = 0.66, p-value = 3.2e-8), TNFα signaling via NF-κB (ES = 0.65, p-value = 8.8e-17) and Epithelial Mesenchymal Transition (ES = 0.64, p-value = 5.7e-16) (**Figure 5A**). These pathways all have some of the largest effect sizes and the lowest p-values. Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Regulation of Immune System Process (ES = 0.56, p-value = 2.15e-19), Inflammatory Response (ES = 0.57, p-value = 1.2e-22), Synapse Organization (ES = 0.63, p-value = 5.7e-18), Synaptic

A



B

Figure 5. Panel plot of CERNO results from the PFC, All Sexes group. A) CERNO + Hallmark, B) CERNO + GOBP.

Signaling (ES = 0.60, p-value = 4.1e-25), Exocytosis (ES = 0.62, p-value = 2.1e-33) and Neurogenesis (ES = 0.58, p-value = 1.9e-21) (**Figure 5B**).

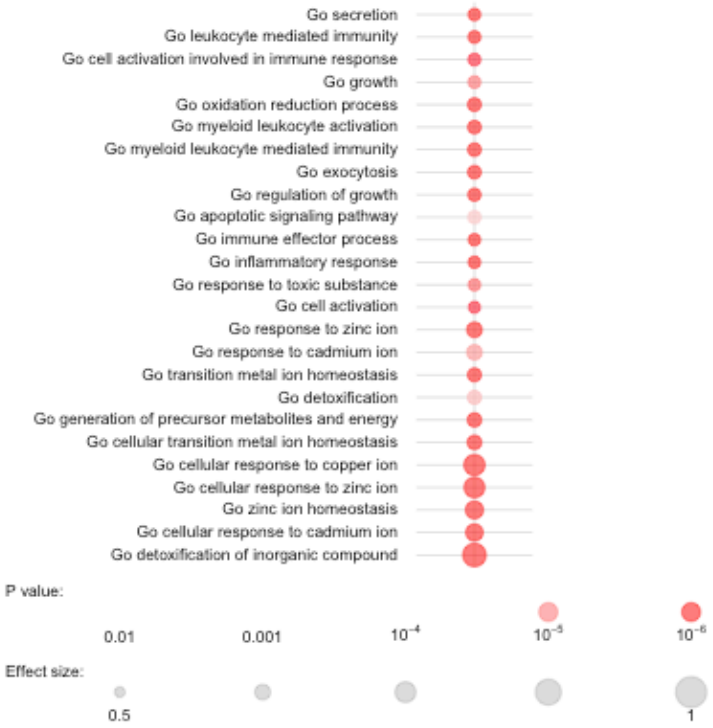
Within the MTG, analyzing the All-Sexes group using CERNO with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Oxidative Phosphorylation (ES = 0.65, p-value = 3.8e-10), Epithelial to Mesenchymal Transition (ES = 0.57, p-value = 2.6e-6), p53 Pathway (ES = 0.61, p-value = 1.2e-6), and Hedgehog Signaling (ES = 0.68, p-value = 2.7e-5) (**see Appendix B, Figure 19B**). Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Neurogenesis (ES = 0.57, p-value = 1.6e-23), Synaptic Signaling (ES = 0.58, p-value = 6.4e-23), Regulation of Vesicle Mediated Transport (ES = 0.61, p-value = 3.5e-18) and Neurotransmitter Transport (ES = 0.60, p-value = 9.9e-17) (**see Appendix B, Figure 21B**).

Within the VC, analyzing the All-Sexes group using CERNO with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Oxidative Phosphorylation (ES = 0.74, p-value = 2.5e-13), TNFα Signaling via NF-κB (ES = 0.65, p-value = 5.6e-16), Inflammatory Response (ES = 0.62, p-value = 1.6e-14), Epithelial to Mesenchymal Transition (ES = 0.65, p-value = 3.9e-15), and Hypoxia (ES = 0.64, p-value = 3.9e-15) (**see Appendix B, Figure 19C**). Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Secretion (ES = 0.59, p-value = 1.6e-34), Exocytosis (ES = 0.62, p-value = 1.1e-30), Synaptic Signaling (ES = 0.60, 6.8e-25) and Myeloid Leukocyte Activation (ES = 0.62, p-value = 1.3e-24) (**see Appendix B, Figure 21C**).

Within the CR, analyzing the All-Sexes group using CERNO with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Oxidative Phosphorylation (ES = 0.70, p-value = 5.0e-8), TNFα Signaling via NF-κB (ES = 0.61, p-value = 1.4e-15), Interferon Gamma Response (ES = 0.57, p-value = 6.7e-6), Glycolysis (ES = 0.58, p-value = 4.5e-5) and Hypoxia (ES = 0.60, p-value = 7.9e-11) (**see Appendix B, Figure 19D**). Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Detoxification of Inorganic Compound (ES = 0.81, p-value = 3.3e-9), Cellular Response to Zinc Ion (ES = 0.75, p-value = 2.9e-9), Zinc Ion Homeostasis (ES = 0.68, p-value = 2.64e-10), and Cellular Response to Cadmium Ion (ES = 0.67, p-value = 2.6e-7) (**see Figure 6A**). Some of the enriched pathways with slightly lower, yet still considerable, effect scores include Electron Transport Chain (ES = 0.66, p-value = 1.9e-5), Cell Activation Involved in Immune Response (ES = 0.55, p-value = 1.5e-6), Inflammatory Response (ES = 0.54, p-value = 7.1e-7), Exocytosis (ES = 0.57, p-value = 1.4e-11).

The CERNO + Hallmark results indicate very similar responses across all brain regions. All regions exhibit preferential enrichment of pathways related to Inflammation, Dedifferentiation, Cell Cycle, and Metabolism. The same enrichment process was repeated for the Female group and Male group across all brain regions. The results indicated similar functional enrichment across all groups and brain regions, however both the CR and MTG brain regions indicated lower enrichment significance overall (**See Appendix B, Figure 19, Figure 20, Figure 21).**

A



B

Figure 6. Panel plot of CERNO + GOBP results from the CR. A) All-Sexes group, B) Female & Male groups.

The CERNO + GOBP results indicate largely similar functional pathway enrichment across all four brain regions, with preferential enrichment of Inflammation, Synaptic Signaling, and Neurogenesis in the PFC; Synaptic Signaling and Neurogenesis in the MTG; Inflammation, Synaptic Signaling and Neurogenesis in the VC; and Inflammation in the CR. The Cerebellum also, interestingly, exhibited slightly elevated enrichment related to Metal Detoxification. This was further investigated in in the Male and Female groups. The detoxification response was seen to be similarly enriched across both sexes (**see Figure 6B**); however, the Males additionally indicated a slightly more significant enrichment of Mitochondrial Metabolism.

fast Gene Set Enrichment Analysis, Multilevel (fGSEA Multilevel)

All *fGSEA* results were ranked by a combination of Normalized Enrichment Score (NES) and a Benjamani-Hochberg adjusted p-value. Within the PFC, analyzing the All-Sexes group using *fGSEA* with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Interferon Gamma Response (NES = 2.76, p-value = 3.6e-20), Inflammatory Response (NES = 2.46, p-value = 3.1e-14), Epithelial to Mesenchymal Transition (NES = 2.36, p-value = 1.1e-12), Oxidative Phosphorylation (NES = -2.80, p-value = 2.3e-21) and TNFα Signaling via NF-κB (NES = 2.29, p-value = 2.7e-12) (**see Figure 7A**). Interestingly, another observed result was increased Estrogen Response and decreased Spermatogenesis. To investigate these results further, a hypergeometric was ran on the genes leading to the enrichment of those specific pathways. The leading genes for the Estrogen Response were also related to inflammation related pathways, the increase of which is consistent with the other results. The leading genes for Spermatogenesis were also related to Cell Cycle and Metabolism, both of which are consistent with the other findings. Using the GOBP gene set

database resulted in the following functional pathways to be some of the most highly enriched terms: Synaptic Signaling (NES = -2.24, p-value = 1.3e-22), Blood Vessel Morphogenesis (NES = 2.39, p-value = 1.1e-23), Inflammatory Response (NES = 2.18, p-value = 8.1e-20) and Cytokine Mediated Signaling Pathway (NES = 2.18, p-value = 1.0e-22) (**see Figure 7B**).

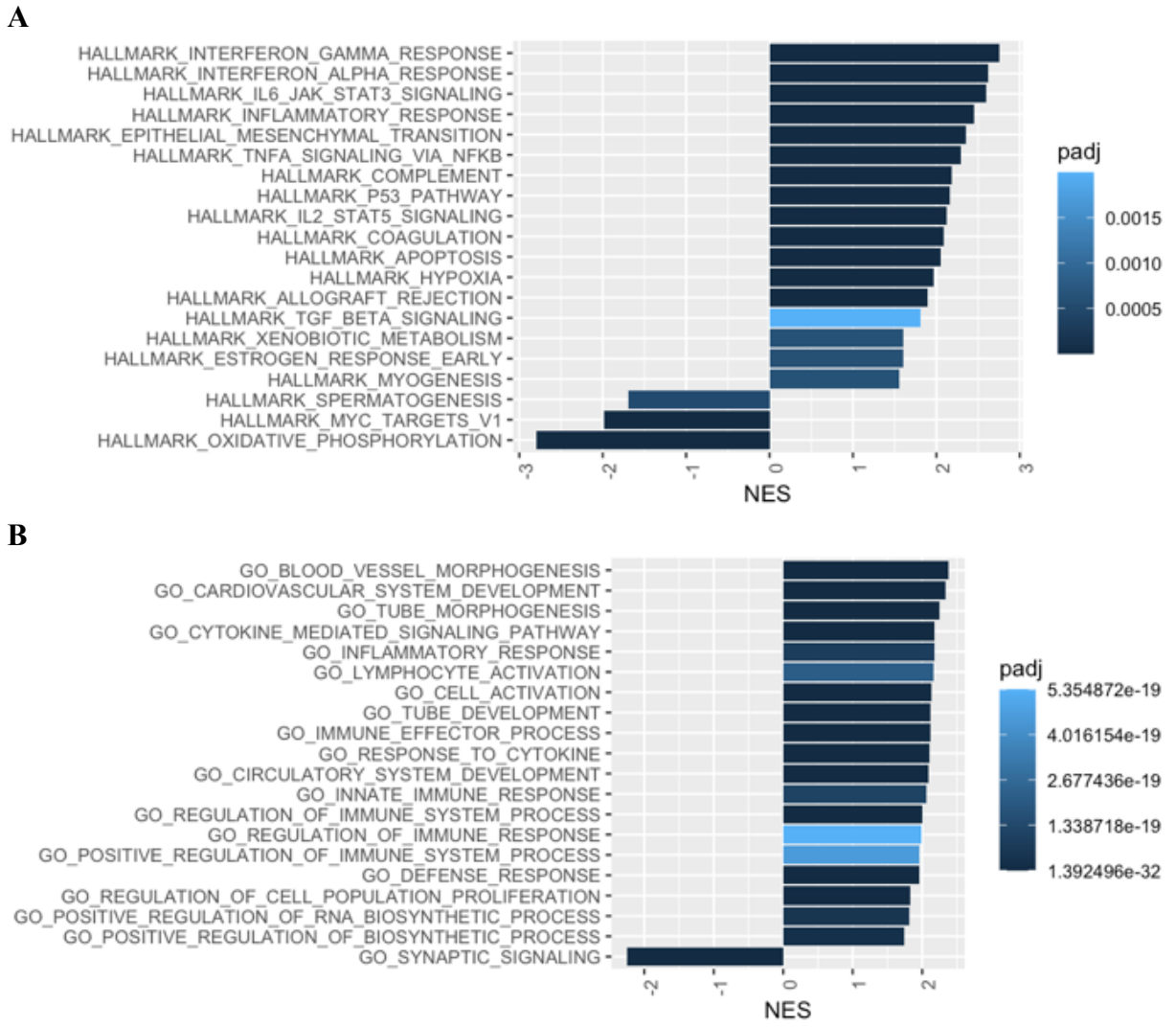


Figure 7. Top fGSEA results for the PFC, All-Sexes group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) fGSEA + Hallmark, B) fGSEA + GOBP.

Within the MTG, analyzing the All-Sexes group using fGSEA with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Interferon Gamma Response (NES = 2.12, p-value = 2.9e-7), TNFα Signaling via NF-κB (NES = 2.03, p-value = 2.5e-6), Epithelial to Mesenchymal Transition (NES = 1.77, p-value = 1.6e-4) and Oxidative Phosphorylation (NES = -2.77, p-value = 8.0e-16) (**see Appendix C, Figure 23B**). Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Blood Vessel Morphogenesis (NES = 2.00, p-value = 3.6e-11), Cell to Cell Signaling (NES = -1.58, p-value = 3.8e-9), Respiratory Electron Transport Chain (NES = -2.72, p-value = 4.2e-10), ATP Synthesis Coupled Electron Transport (NES = -2.80, p-value = 2.7e-10) and Synaptic Signaling (NES = -2.31, p-value = 7.9e-21) (**see Appendix D, Figure 26B).**

Within the VC, analyzing the All-Sexes group using fGSEA with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Interferon Gamma Response (NES = 2.83, p-value = 1.2e-19), Inflammatory Response (NES = 2.81, p-value = 8.7e-19), Interferon Alpha Response (NES = 2.74, p-value = 6.4e-12), Oxidative Phosphorylation (NES = -2.93, p-value = 1.5e-22) and Epithelial to Mesenchymal Transition (NES = 2.19, p-value = 6.0e-9) (**see Appendix C, Figure 23C**). Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Blood Vessel Morphogenesis (NES = 2.56, p-value = 1.2e-27), Myeloid Leukocyte Activation (NES = 2.27, p-value = 5.3e-20), Inflammatory Response (NES = 2.24, p-value = 4.6e-21) and Cytokine Mediated Signaling Pathway (NES = 2.19, p-value = 4.6e-21) (**see Figure 8A**). Interestingly, this was the only brain region without the reduced Metabolism pathways in the highest enriched terms. These results are mirrored in the enrichment of the Female group (**see**

Figure 8B) and largely similar to those of the Male group (**see Figure 8C**). However, it appears that the Female group is driving the overall enrichment for the VC brain region.

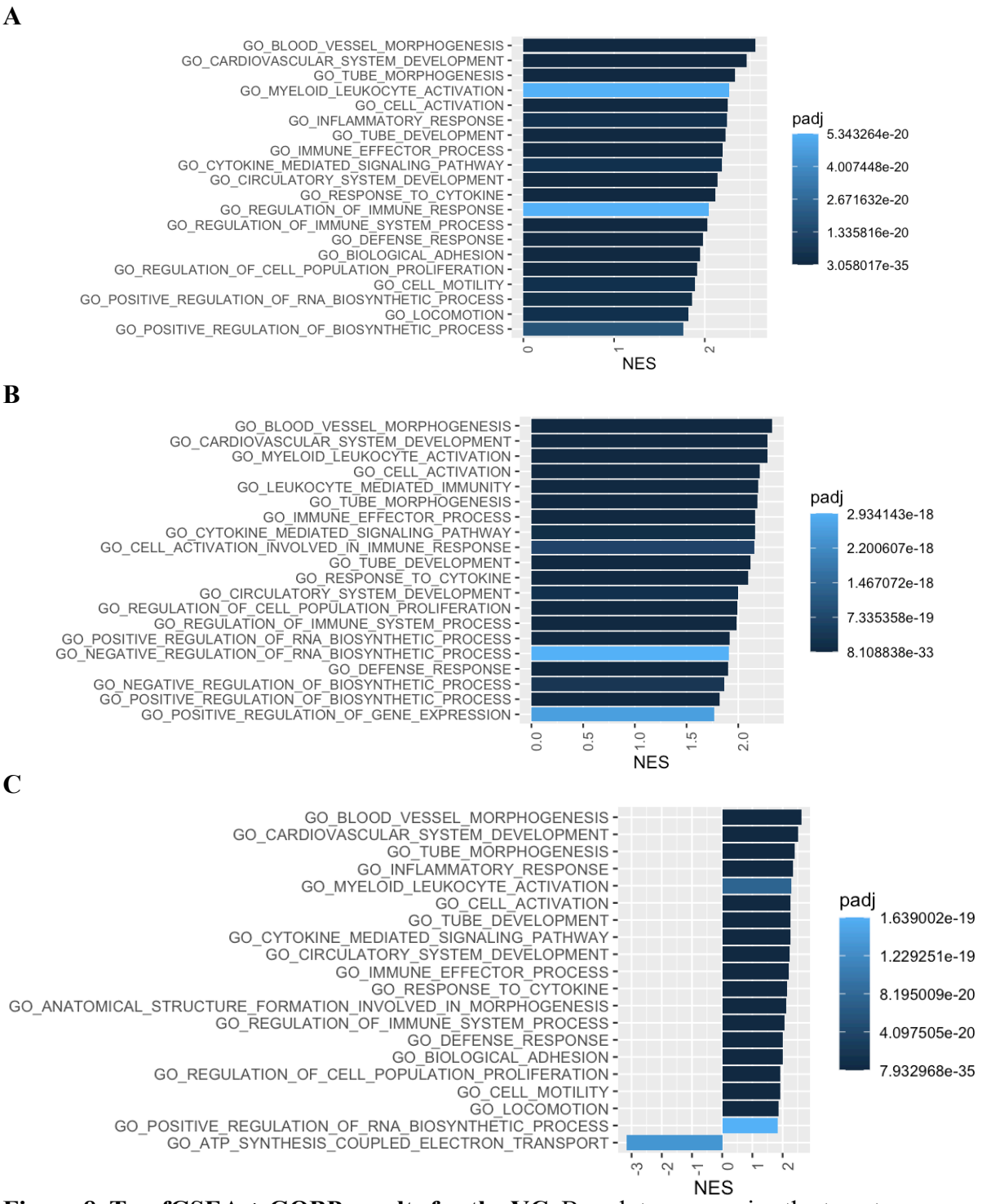


Figure 8. Top fGSEA + GOBP results for the VC. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) All-Sexes group, B) Females, C) Males.

Within the CR, analyzing the All-Sexes group using fGSEA with the Hallmark database resulted in the following functional pathways to be some of the most highly enriched terms: Inflammatory Response (NES = 2.30, p-value = 4.96e-12), Interferon Gamma Response (NES = 2.30, p-value = 8.7e-11), G2M Checkpoint (NES = 1.76, p-value = 1.1e-4), Oxidative Phosphorylation (NES = -2.69, p-value = 3.9e-20) and Epithelial to Mesenchymal Transition (NES = 1.72, p-value = 1.0e-4) (**see Appendix C, Figure 23D**). Using the GOBP gene set database resulted in the following functional pathways to be some of the most highly enriched terms: Blood Vessel Morphogenesis (NES = 2.12, p-value = 1.5e-15), Cellular Respiration (NES = -2.58, p-value = 1.6e-16), Oxidative Phosphorylation (NES = -2.79, p-value = 2.6e-15) and Regulation of Immune System Process (NES = 1.74, p-value = 2.0e-16) (**see Appendix D, Figure 26D**).

The fGSEA + Hallmark gene set enrichment results of the All-Sexes group indicate similar functional pathway enrichment across all four brain regions, with all regions exhibiting similar enriched pathways related to increased Inflammatory Response, reduced Oxidative Phosphorylation and increased Dedifferentiation. Additional pathways with lower, yet still considerable, statistical significance indicated cell cycle process enrichment consistently across brain regions. Enrichment of the Female and Male groups indicated similar results, with all brain regions indicating decreased metabolism, increased dedifferentiation and cell cycle activity (**see Appendix C, Figure 23, Figure 24, Figure 25).**

The fGSEA + GOBP gene set enrichment results indicate similar functional pathway enrichment across all four brain regions, with all regions exhibiting similar enriched pathways related to increased Inflammatory Response, decreased Synaptic Signaling, increased Blood Vessel and Tube Development and decreased Metabolism. The results suggest slightly greater

statistically significant preferential enrichment of Synaptic Signaling in the PFC and MTG and Metabolism in the MTG and VC. It should be noted that the genes related to the enrichment of vasculature development related pathways, seen in the aforementioned results, are also related to Epithelial to Mesenchymal Transition. For this reason, the enrichment of these terms is considered to also be indicative of a Dedifferentiation functional process.

Enrichment analysis of the Female group of each brain region indicated a similar trend of results as that of the All-Sexes group: increased Blood Vessel Development and Inflammation response across all regions, with the PFC and MTG indicating preferential decreased enrichment of Synaptic Signaling (**See Appendix D, Figure 27).** Enrichment analysis of the Male group of each brain region also indicated a similar trend of results: pathways related to Blood Vessel Development and Inflammation were positively enriched across all regions, pathways related to Synaptic Signaling were negatively enriched in the PFC and MTG and pathways related to Metabolism were negatively enriched in the MTG and VC (**See Appendix D, Figure 28).**

It should also be noted that similar to the CERNO results, the MTG and CR displayed reduced overall enrichment significance, compared to the PFC and VC. The results generated from CERNO and fGSEA were largely similar, indicating enrichment of inflammation, dedifferentiation, cell cycle and metabolism. One of the fundamental differences between the results is that the fGSEA results offer insight into enrichment directionality: increased enrichment of inflammation, dedifferentiation and cell cycle reentry, and decreased enrichment of metabolism and synaptic signaling. Using both methods allows for more informed analysis of the data and the similarities of the results from both methods allows for stronger confidence in the enrichment results.

Transcription Factor Enrichment

Transcription factor (TF) enrichment allows for identification of active TFs from TF target gene expression data in order to allow inferences into functional analysis (Alvarez, 2020).

Transcription Factor Enrichment Methods

fast Gene Set Enrichment Analysis, Multilevel (fGSEA Multilevel)

The fGSEA enrichment test, as previously described (see page 9), can also be used for TF enrichment when paired with appropriate databases. For this test, the t-value ranking was used as the input. The Encode Chea Consensus gene set database, generated from the ENCODE and Chea data, was used for TF enrichment (Lachmann, et al., 2010; Davis, et al., 2018; Chen, et al., 2013; Kuleshov, et al., 2016).

Discriminant Regulon Expression Analysis (DoRothEA)

This enrichment test method couples DoRothEA, a gene set resource containing curated regulons, and VIPER, a statistical method used for enrichment analysis of regulons (Alvarez, 2020). DoRothEA's regulons are defined as TFs and their transcriptional targets, which have been curated using evidence including literature, ChIP-seq peaks, TF binding site motifs and inference from gene expression (Garcia-Alonso, Holland, Ibrahim, Turei, & Saez-Rodriguez, 2019). Each regulon is bidirectionally ranked by confidence level, A (highest) – E (lowest), indicating repression or activation as determined by the amount of supporting evidence.

For this test, t-value ranking was used as the input. The regulons classified with a confidence level of "C" were used, which is the mid-level confidence ranking resulting from

literature curated and/or ChIP-seq peak evidence, with additional levels of evidence (Garcia-Alonso, Holland, Ibrahim, Turei, & Saez-Rodriguez, 2019).

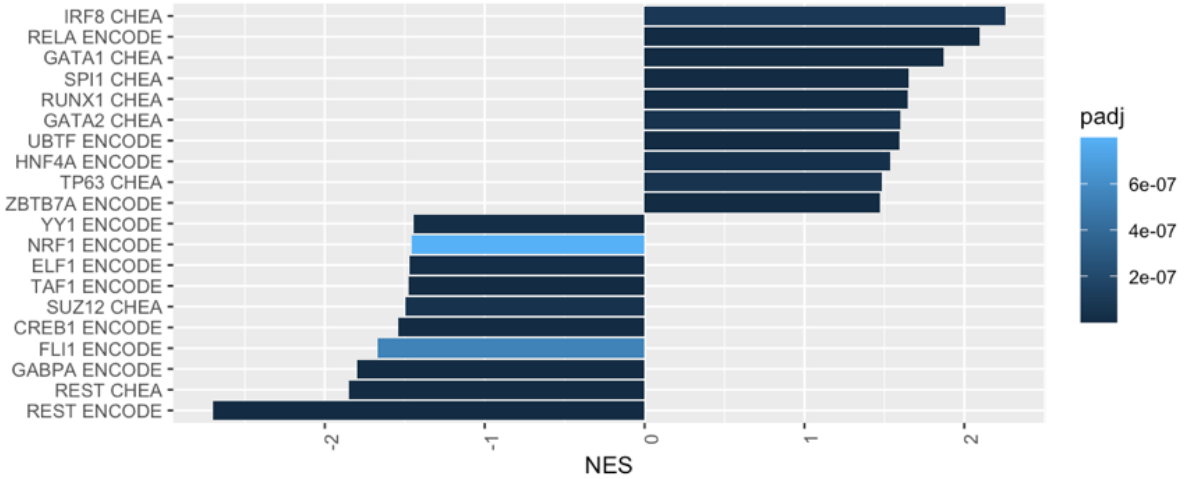


Figure 9. Top regulons of fGSEA + ECC results for PFC, All Sexes group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value.

Transcription Factor Enrichment Results

fast Gene Set Enrichment Analysis, Multilevel (fGSEA Multilevel)

Within the PFC, analyzing the All-Sexes group using fGSEA with the ECC database resulted in the following TFs to be some of the most highly enriched terms: IRF8 (NES = 2.26, p-value = 9.3e-8), RELA (NES = 2.09, p-value = 3.2e-12), GATA1 (NES = 1.87, p-value = 1.8e-13), SPI1 (NES = 1.65, p-value = 9.3e-11), GABPA (NES = -1.80, p-value = 1.2e-19) and REST (NES = -2.70, p-value = 2.6e-27), (**see Figure 9**).

Within the MTG, analyzing the All-Sexes group using fGSEA with the ECC database resulted in the following TFs to be some of the most highly enriched terms: GATA1 (NES = 1.75, p-value = 5.9e-9), FOXA1 (NES = 1.73, p-value = 5.2e-4), SUZ12 (NES = -1.52, p-value = 3.3e-8), TAF1 (NES = -1.52, p-value = 5.4e-14), REST (NES = -2.87, p-value = 2.1e-27) (**see Appendix E, Figure 29B**).

Within the VC, analyzing the All-Sexes group using fGSEA with the ECC database resulted in the following TFs to be some of the most highly enriched terms: RELA (NES = 2.28, p-value = 1.8e-15), GATA1 (NES = 1.98, p-value = 4.9e-16), SOX2 (NES = 1.68, p-value = 5.0e-9), GABPA (NES = -2.11, p-value = 1.6e-35) and REST (NES = -2.71, p-value = 3.8e-25) (**see Appendix E, Figure 29C**).

Within the CR, analyzing the All-Sexes group using fGSEA with the ECC database resulted in the following TFs to be some of the most highly enriched terms: GATA1 (NES = 1.71, p-value = 2.1e-10), SOX2 (NES = 1.63, p-value = 5.7e-9), MYC (NES = -1.78, p-value = 1.5e-15), GABPA (NES = -1.99, p-value = 2.5e-31) and REST (NES = -2.15, p-value = 2.6e-13) (**see Appendix E, Figure 29D)**.

All brain regions exhibited similar TF enrichment trends, with an emphasis on inflammation and cell cycle related TFs. The enrichment of proinflammatory related TFs, such as RELA (Bettelli, Dastrange, & Oukka, 2005) and IRF8 (Agod, et al., 2018), in the PFC and VC support the previous findings of an inflammatory immune response. The enrichment of TFs related to cell cycle activity, such as CREB1 (Desdouets, et al., 1995) and MYC (García-Gutiérrez, Delgado, & León, 2019), observed across all brain regions, supports the previous findings of gene set enrichment related to cell cycle activity. REST (Lunyak, et al., 2002), a canonical repressor of neuronal lineage, was decreased in all brain regions. Additionally, negative enrichment of NRF1 (Gopalakrishnan & Scarpulla, 1995) in the VC and PFC, and GABPA (Yang, Drumea, Mott, Wang, & Rosmarin, 2014) in all regions, suggest reduced metabolic activity in all brain regions, again supporting the suppression of metabolism pathways found in the gene enrichment results previously described.

TF enrichment performed on the Female and Male groups indicated similar trends of results as the All-Sexes group for all brain regions: increased enrichment of cell cycle related TFs, proinflammatory TFs, suppression of neurogenesis repressor REST and suppression of metabolic related TFs (see **Appendix E, Figure 30, Figure 31**). Also similar to the previous findings, the MTG had lower enrichment significance in the All-Sexes and Male groups, compared to the other regions.

Discriminant Regulon Expression Analysis (DoRothEA)

Within the PFC, analyzing the All-Sexes group using DoRothEA resulted in the following regulons to be some of the most highly enriched terms: STAT1 (NES = 10.94, p-value = 7.3e-28), SPI1 (NES = 8.63, p-value = 6.2e-18), NF-κB1 (NES = 8.36, p-value = 6.1e-17), ZEB1 (NES = -5.78, p-value = 7.5e-9) and ZNF263 (NES = -5.70, p-value = 1.2e-8) (**see Figure 10**).

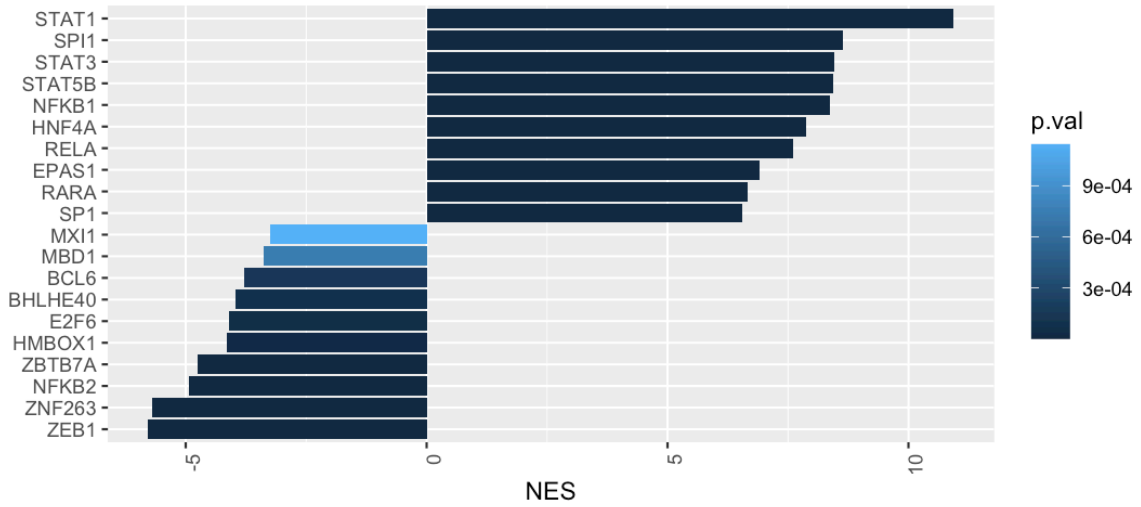


Figure 10. Top 20 regulons of DoRothEA results for PFC, All Sexes. Bar plots summarize the regulons ordered by NES. Bar color indicates p-value.

Within the MTG, analyzing the All-Sexes group using DoRothEA resulted in the following regulons to be some of the most highly enriched terms: RELA (NES = 6.91, p-value =

4.9e-12), HNF4A (NES = 6.12, p-value = 9.2e-10), STAT1 (NES = 6.10, p-value = 1.1e-9), BCL6 (NES = -3.96, p-value = 7.6e-5) and ZEB1 (NES = -4.60, p-value = 4.2e-6) (**see Appendix F, Figure 32B).**

Within the VC, analyzing the All-Sexes group using DoRothEA resulted in the following regulons to be some of the most highly enriched terms: STAT1 (NES = 10.10, p-value = 3.6e-24), STAT3 (NES = 8.66, p-value = 4.7e-18), NF-κB1 (NES = 8.26, p-value = 1.4e-16), RELA (NES = 8.14, p-value = 3.9e-16) and ZNF263 (NES = -5.80, p-value = 6.7e-9) (**see Appendix F, Figure 32C**).

Within the CR, analyzing the All-Sexes group using DoRothEA resulted in the following regulons to be some of the most highly enriched terms: STAT3 (NES = 8.28, p-value = 1.2e-16), STAT1 (NES = 8.00, p-value = 1.2e-15), SPI1 (NES = 7.34, p-value = 2.2e-13), NF-κB1 (NES = 6.73, p-value = 1.7e-11) and NF-κB2 (NES = -5.49, p-value = 4.0e-8) (**see Appendix F, Figure 32D**).

The TF enrichment results from DoRothEA suggest immune and inflammation activity, as indicated by the increase of proinflammatory markers RELA (Bettelli, Dastrange, & Oukka, 2005), STAT1 (Chen, et al., 2017) and STAT3 (Ma, Huang, Wang, & Xin, 2017). This trend was observed across all brain regions. E2F6, a regulator of genes required for cell cycle entry (Cartwright, Müller, Wagener, Holm, & Helin, 1998), was suppressed in all brain regions. Additionally, ZEB1, an activator of neurogenesis (Wellner, Schubert, Burk, & Schmalhofer, 2009), was suppressed in all brain regions. ZEB1 is also believed to be related to dedifferentation (Wellner, Schubert, Burk, & Schmalhofer, 2009). These results are consistent across all brain regions. These results were also very similar across the Female and Male groups across all brain regions (**see Appendix F, Figure 33, Figure 34**).

ENDOTYPE ANALYSIS

Endotype Selection

The gene set and TF enrichment results were used to select five main endotypes, pathophysiological mechanisms, that were well represented across all brain regions: Inflammation, Dedifferentiation, Cell Cycle, Metabolism, and Synaptic Signaling. These endotypes were studied to gain insight into potential mechanisms of AD across the four brain regions of interest.

Custom gene sets were then created for each endotype, by taking the union of gene sets of GOBP and ECC that were highly represented across the results (see **Table 4**). For the gene set for Dedifferentiation, dbEMT (V1.0) was incorporated to include additional core EMT genes (Zhao, Kong, Liu, & Qu, 2015) (Zhao D. M., 2020). For any gene initially listed in multiple endotype gene sets, the gene was removed from all but one endotype.

Table 4. Custom Gene Sets for Endotypes

Endotype	***Gene Sets Utilized***	***Size***
Inflammation	GO Inflammatory Response Hallmark Inflammatory Response	839
Dedifferentiation	GO Epithelial to Mesenchymal Transition Hallmark Epithelial to Mesenchymal Transition dbEMT (V1)	592
Cell Cycle	GO Cell Cycle G1S Phase Transition Hallmark E2F Targets	456
Metabolism	GO Oxidative Phosphorylation GO Mitochondrial Respiratory Chain Complex Assembly Go Mitochondrial Electron Transport NADH to Ubiquinone GO Aerobic Respiration Hallmark Oxidative Phosphorylation	335
Synaptic Signaling	GO Synaptic Signaling	713

Protein-Protein Interaction (PPI) Networks

Each endotype gene set was then used to create a PPI network via STRINGdb (Szklarczyk, et al., 2019). The full networks were generated where edges indicate physical and functional protein associations based on experimental and database active interaction sources, with a confidence of at least 0.9. The networks were then exported to Cytoscape (Shannon, et al., 2003) where the TF

to target gene interactions were incorporated.

A hypergeometric distribution, via EnrichR, was then used for each endotype gene set to verify the TFs that would most likely target the custom gene set (Chen, et al., 2013; Kuleshov, et al., 2016). The resulting TFs were compared against the TF enrichment results from fGSEA and DoRothEA and used to select three top TFs for each brain region and endotype (see **Table 5**). TF to target gene interactions were determined using the gene sets of DoRothEA and ECC. For any TF of interest, the targets were identified and intersected with the endotype of focus. This list was then utilized to form the TF to gene target edges of the PPI networks.

When constructing the PPI Networks, the TF-gene and gene-gene interactions were identified with purple and grey lines, respectively. The node colors and sizes were determined by the effect score and the absolute value of the confect score, respectively, from the *topconfects* differential expression results.

Endotype Analysis Results

The PPI networks generated for the Cell Cycle endotype suggest that the G1S phase of Cell Cycle is more affected in the PFC and VC brain regions of AD patients, compared to the MTG and CR brain regions, as indicated by the increased number of larger and more deeply

colored nodes in both the PFC and VC, corresponding to confect and effect scores of each gene (see **Figure 11**). CDK2 was upregulated in the PFC and VC, regulates the G1S phase transition of cell cycle (Cyclin-dependent kinase 2, 2020), whereas NME1, a regulator of nucleoside triphosphates (Fan, Beresford, Oh, Zhang, & Lieberman, 2003), was downregulated in the MTG, VC and CR.

Table 5. TF selected for each endotype, per brain region for the All-Sexes group.

Endotype	***Region***	***All-Sexes***
Cell Cycle	PFC	E2F6, CREB1, ATF2
	MTG	E2F6, CREB1, ATF2
	VC	E2F6, CREB1, ATF2
	CR	MAX, MYC, CREB1
Inflammation	PFC	STAT1, GATA1, RELA
	MTG	RELA, GATA1, RUNX1
	VC	STAT1, GATA1, RELA
	CR	STAT3, GATA1, RUNX1
Dedifferentiation	PFC	SOX2, TCF2, SALL4
	MTG	HNF4A, TFAP2C, AR
	VC	SOX, TCF3, SALL4
	CR	SOX, NFE2L2, SALL4
Synaptic Signaling	PFC	REST, ZNF263, SUZ12
	MTG	REST, SUZ12, EZH2,
	VC	REST, TP53, EZH2
	CR	REST, ZNF263, TP53
Metabolism	PFC	GABPA, NRF1, TAF1
	MTG	TAF1, YY1, GABPA
	VC	GABPA, NRF1, TAF1
	CR	GABPA, NRF1, TAF1

The PPI networks generated for the Inflammation endotype suggest that the inflammatory response is more highly activated in the PFC and VC, and least affected in the MTG (see **Figure 12**). S100A8 and S100A9, proinflammatory genes capable of triggering an inflammatory cascade (Ryckman, Vandal, Rouleau, Talbot, & Tessier, 2003), were two of the most positively enriched genes across the PFC, VC and MTG. CRH, a regulator of neuroendocrine response to stress

(CRH corticotropin releasing hormone, 2020), was a commonly negatively enriched gene across the PFC and VC. The MTG appears to be minimally affected compared to the other regions.

The PPI networks generated for the Dedifferentiation endotype suggest that the PFC and VC brain regions are more highly affected by Dedifferentiation than the MTG or CR in AD patients (see **Figure 13**). SCG2, a regulator of the genesis of secretory granules (Hotta, Hosaka, Tanabe, & Takeuchi, 2009) is repressed in the PFC and VC; and PAK1, a regulator of microtubule biogenesis (Sells, et al., 1997), is repressed in the PFC, VC and CR. GJA1, a key regulator of gap junction communication (Negoro, 2012), is activated in both the VC and CR.

The PPI networks generated for the Synaptic Signaling endotype suggest that the PFC and VC brain regions experience greater inactivation of synaptic signaling in AD patients, than do the MTG or CR brain regions (see **Figure 14**). In a similar trend as previously reported, the MTG and CR appear to be much less affected than the PFC or VC. In the PFC and VC, GFAP, an astrocyte specific marker (Reeves, Helman, Allison, & Israel, 1989), is activated, whereas the majority of other genes are suppressed. One such suppressed gene is SST, an inhibitor of pituitary hormones (Luque & Kineman, 2018), which is suppressed in both the PFC and VC.

The PPI networks generated for the Metabolism endotype suggest that the PFC and VC brain regions are more affected than the MTG and CR brain regions (see **Figure 15**). MDH1, a gene encoding for an enzyme critical for metabolic pathways such as the Krebs Cycle (NCBI, 2020), was strongly inhibited in the PFC and VC. CASP7, a proapoptotic gene (Chai, et al., 2001), was strongly induced in the PFC, VC and CR.

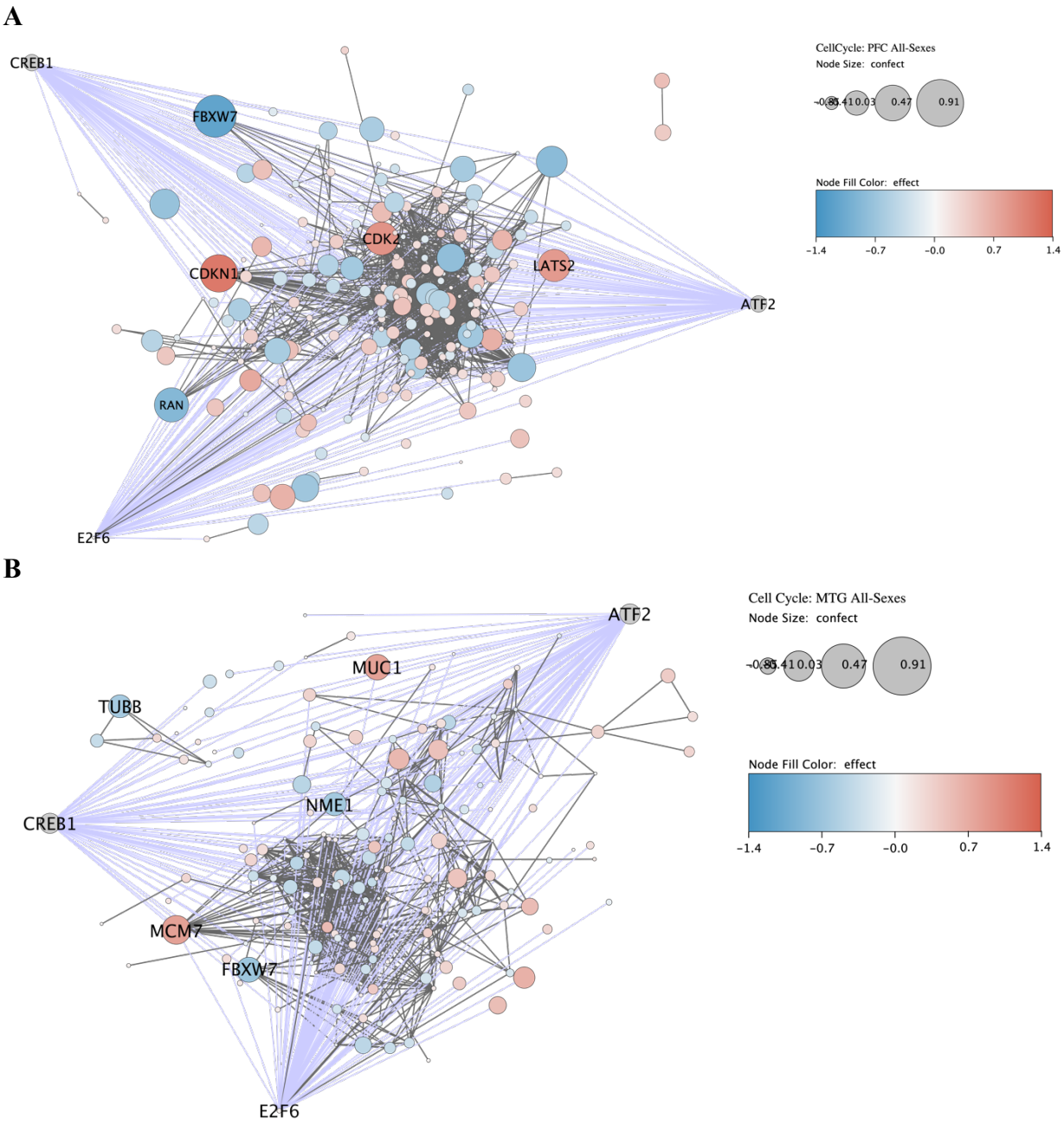
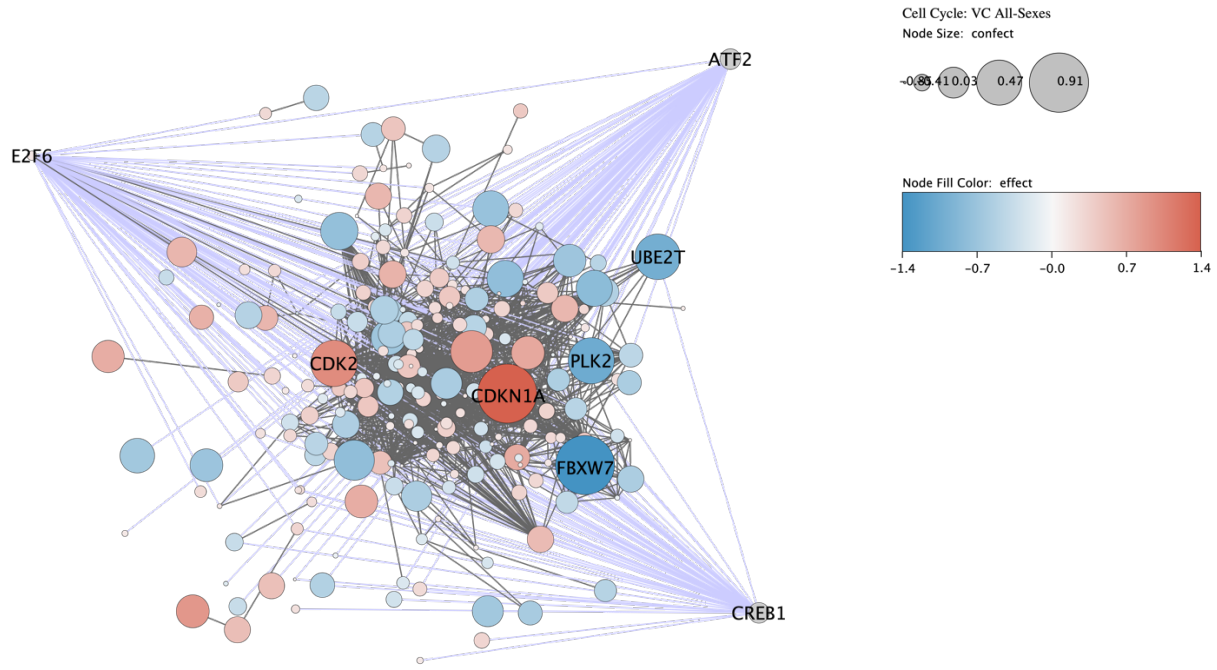


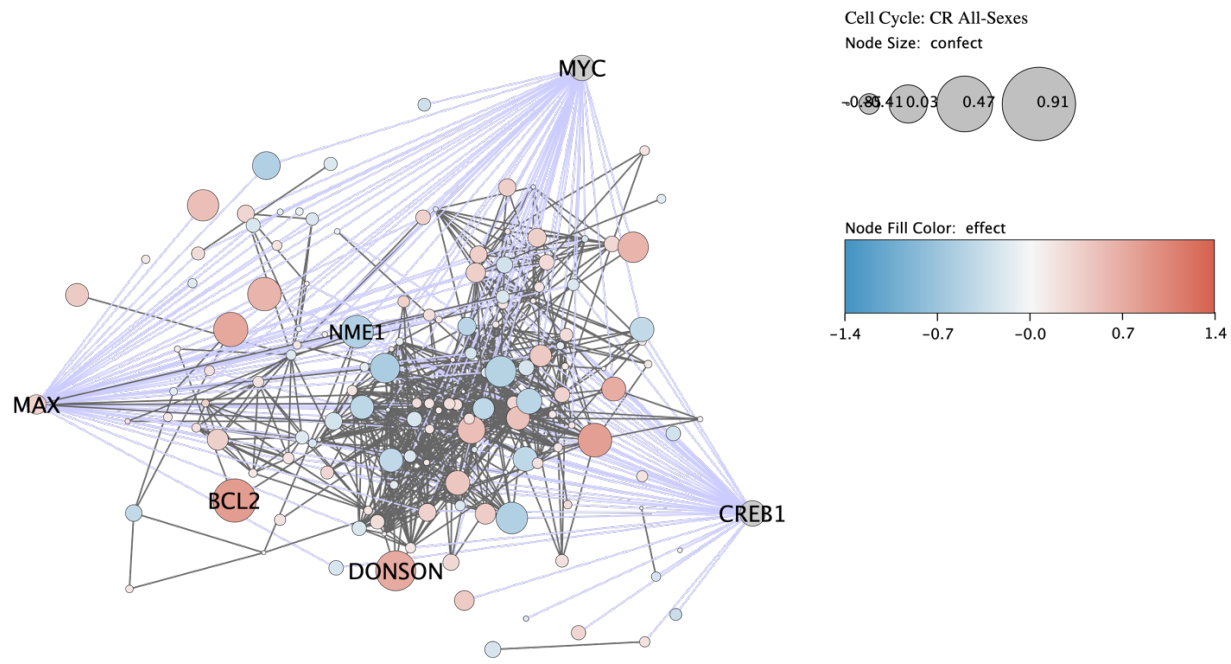
Figure 11. Cell Cycle Endotype PPI Network. A) PFC, B) MTG, C) VC, D) CR. Node size indicates differential expression confect score. Node color indicates differential expression effect. Grey nodes indicate TF target genes absent in the network. Purple lines represent TF-gene edges.

Figure 11 continued.

C



D



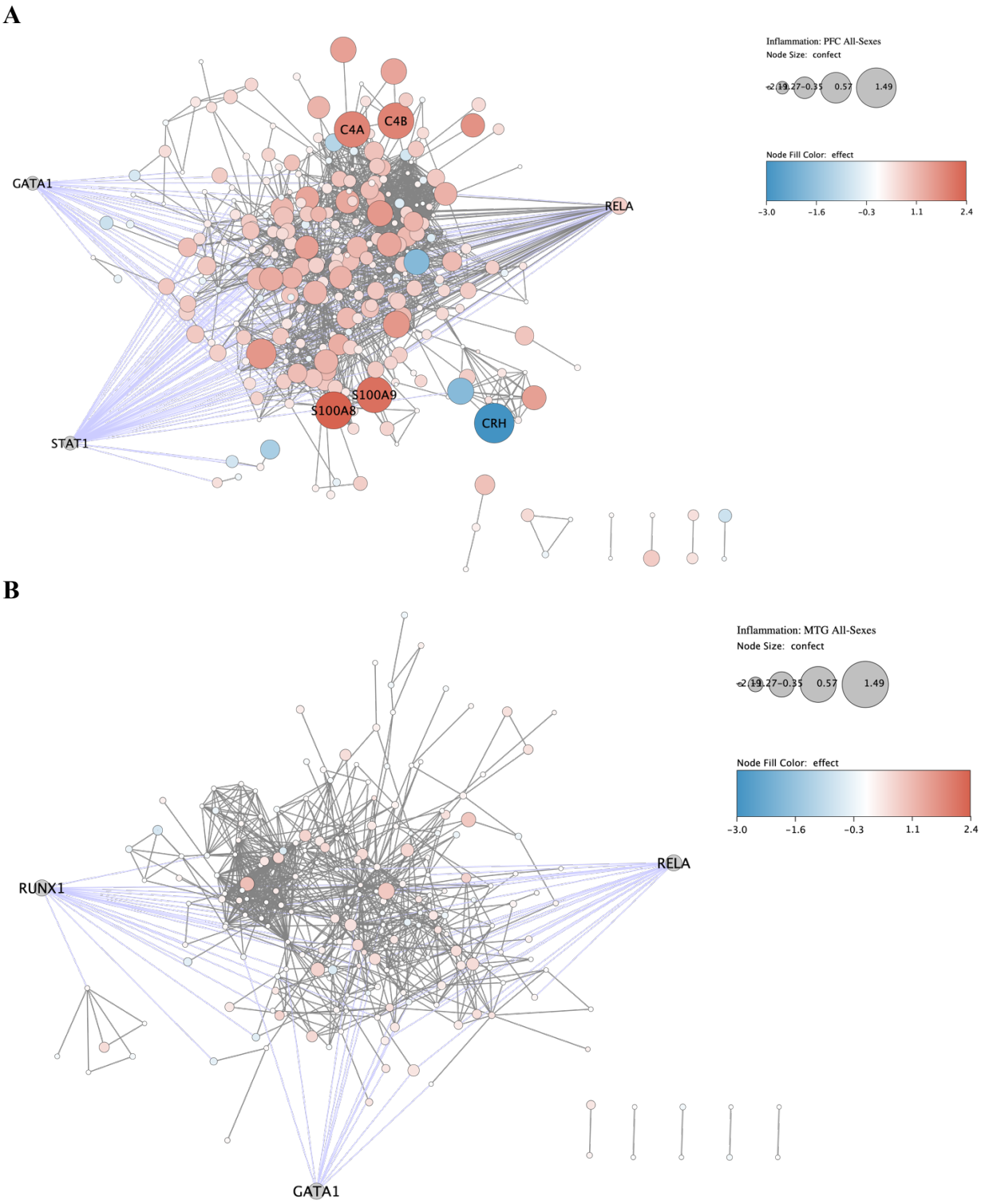
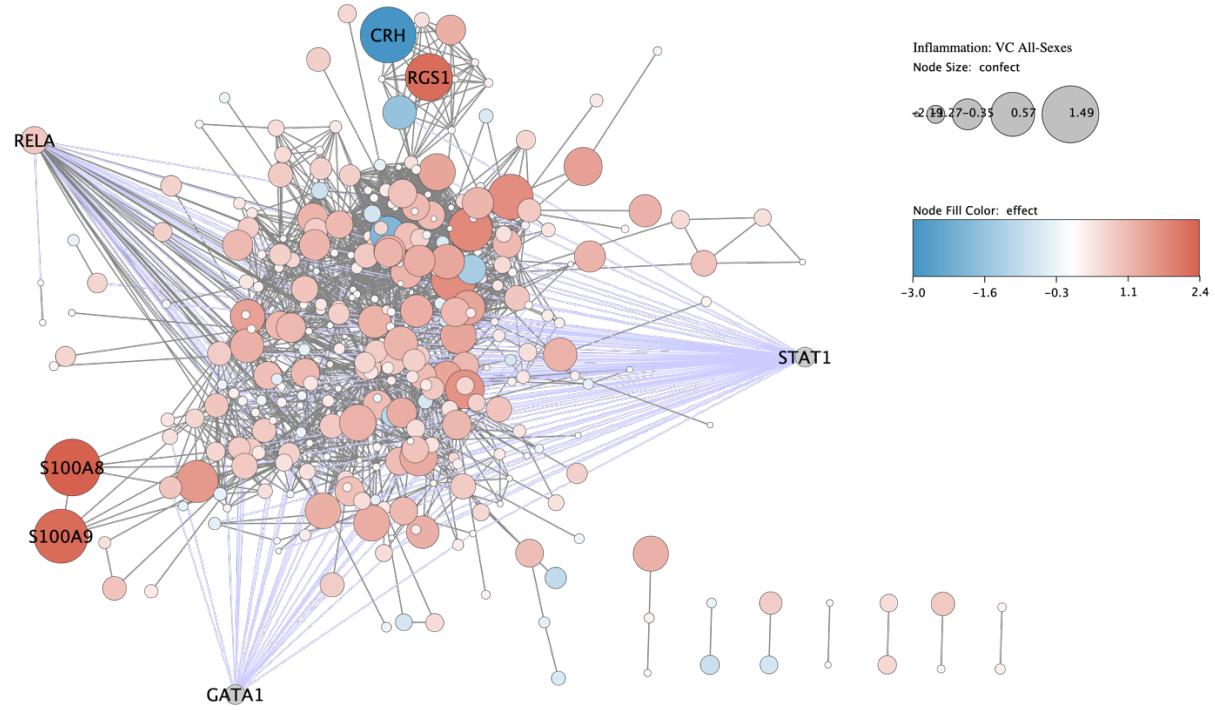


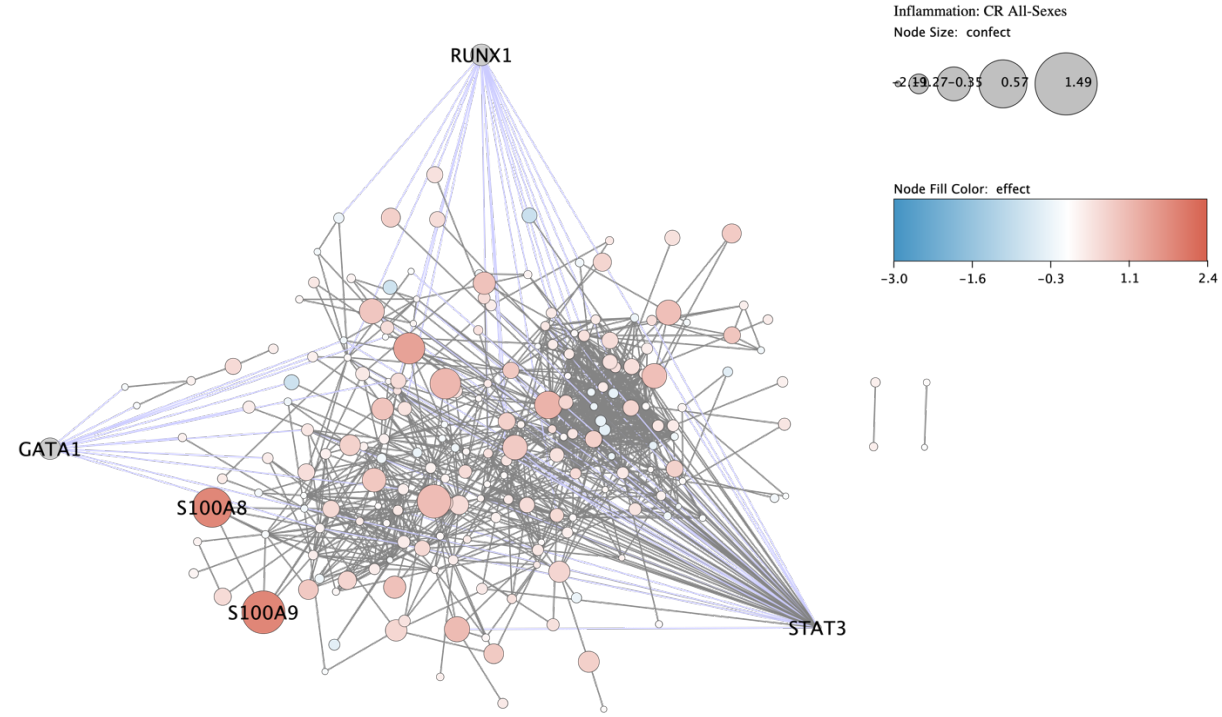
Figure 12. Inflammation Endotype PPI Network. A) PFC, B) MTG, C) VC, D) CR. Node size indicates differential expression confect score. Node color indicates differential expression effect. Grey nodes indicate TF target genes absent in the network. Purple lines represent TF-gene edges.

Figure 12 continued.

C



D



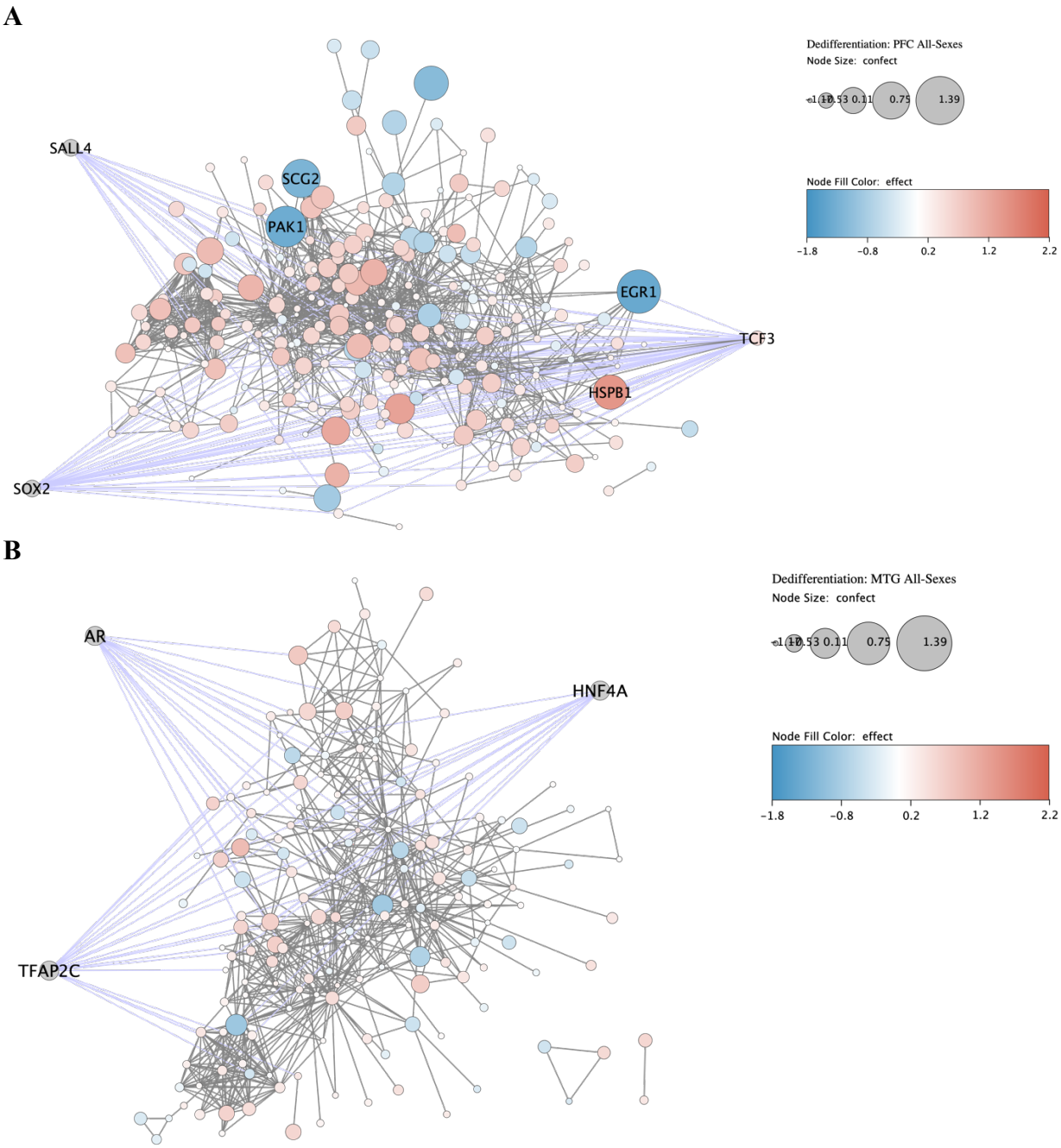
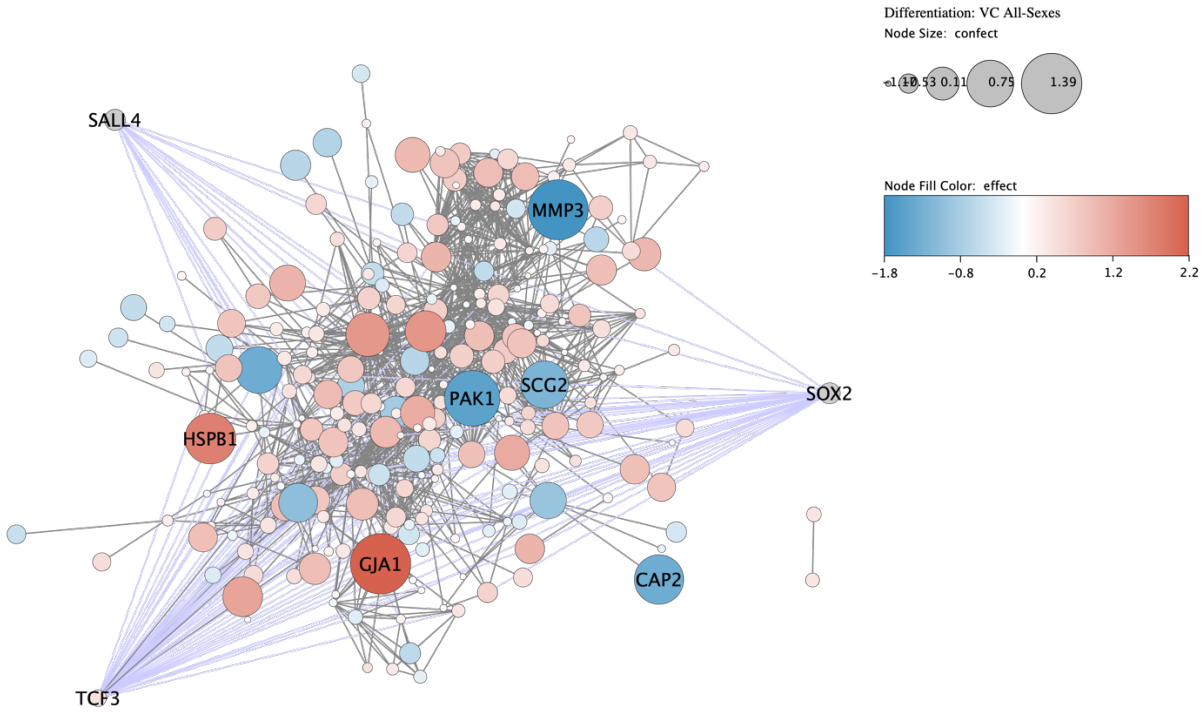


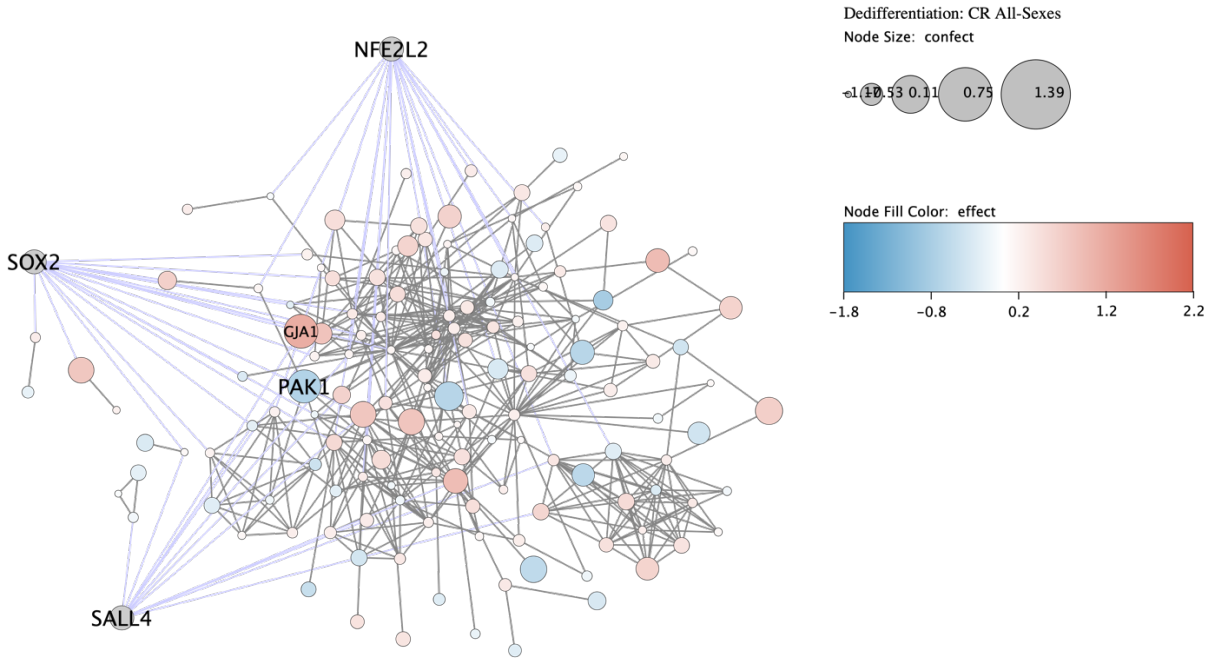
Figure 13. Dedifferentiation Endotype PPI Network. A) PFC, B) MTG, C) VC, D) CR. Node size indicates differential expression confect score. Node color indicates differential expression effect. Grey nodes indicate TF target genes absent in the network. Purple lines represent TF-gene edges.

Figure 13 continued.

C



D



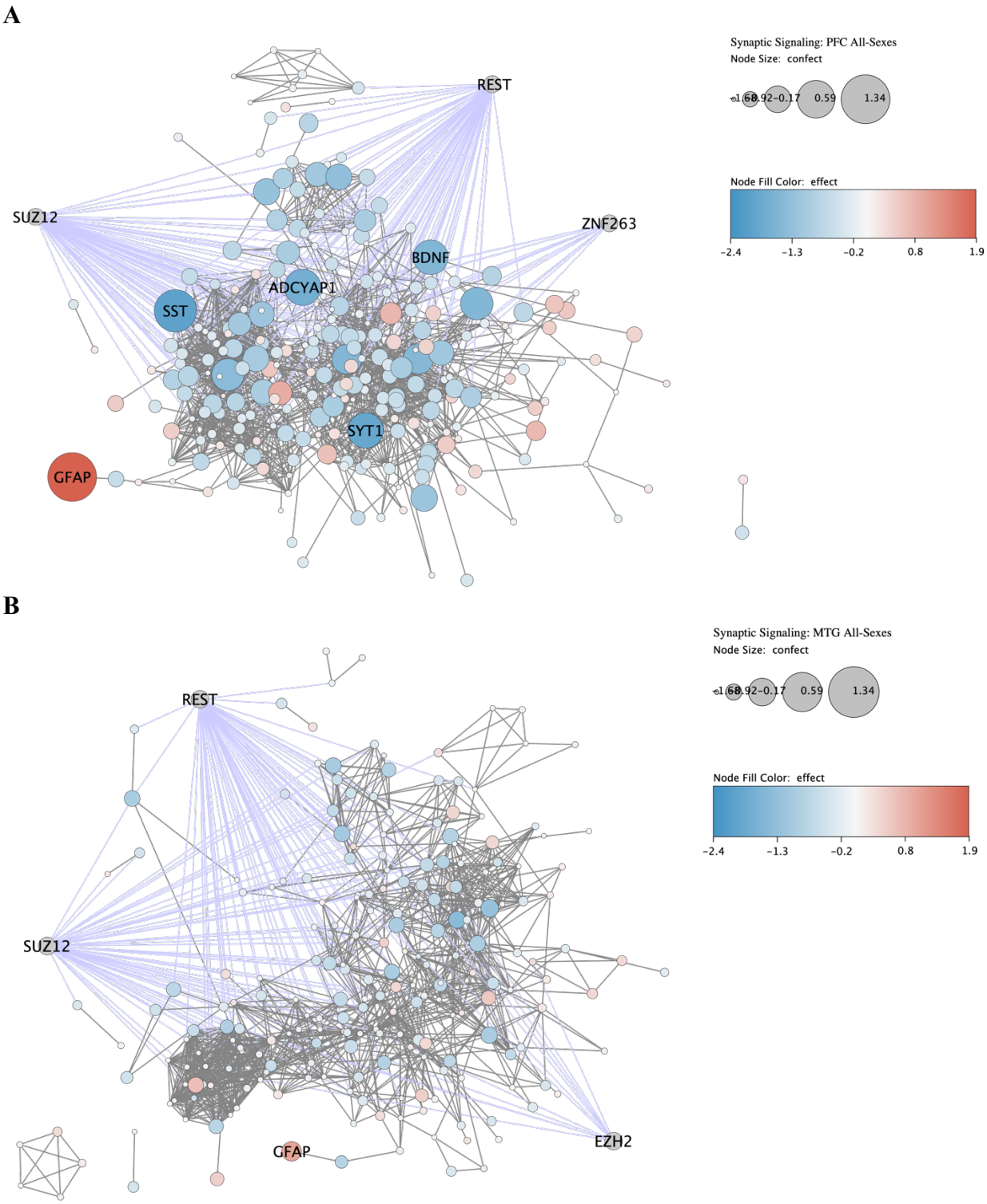
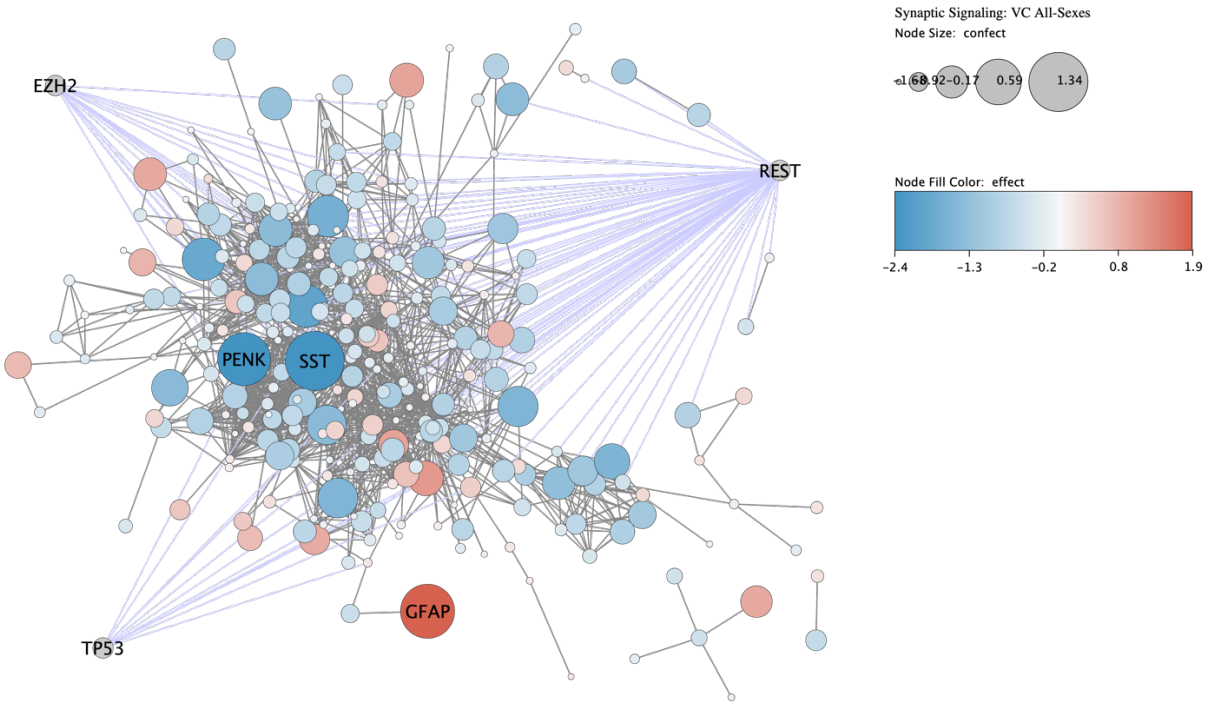


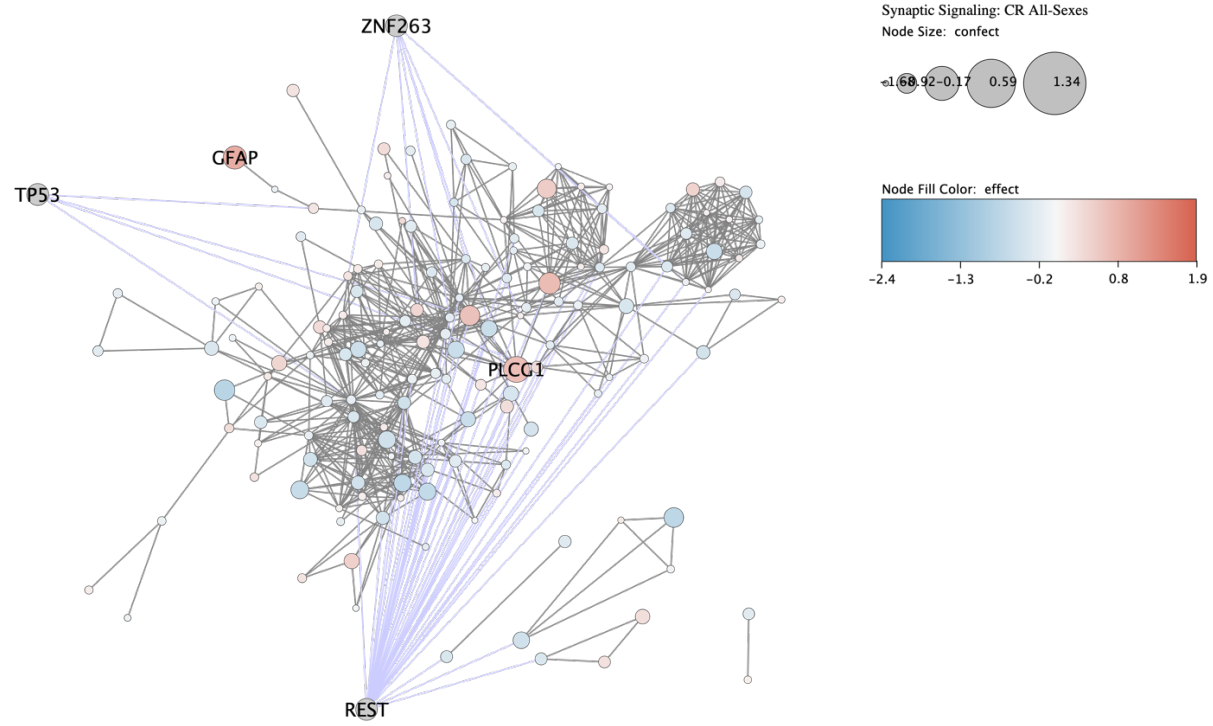
Figure 14. Synaptic Signaling Endotype PPI Network. A) PFC, B) MTG, C) VC, D) CR. Node size indicates differential expression confect score. Node color indicates differential expression effect. Grey nodes indicate TF target genes absent in the network. Purple lines represent TF-gene edges.

Figure 14 continued.

C



D



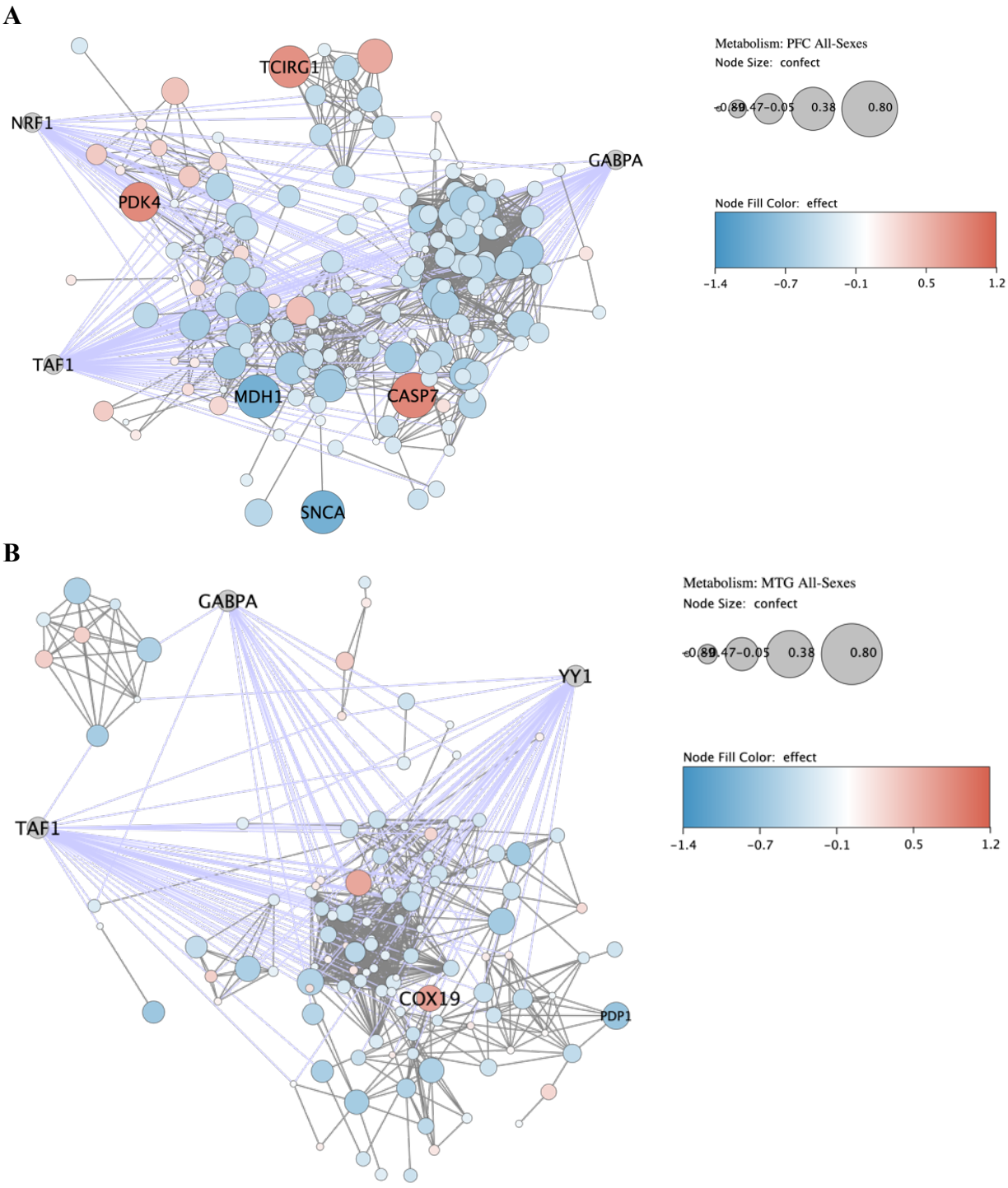
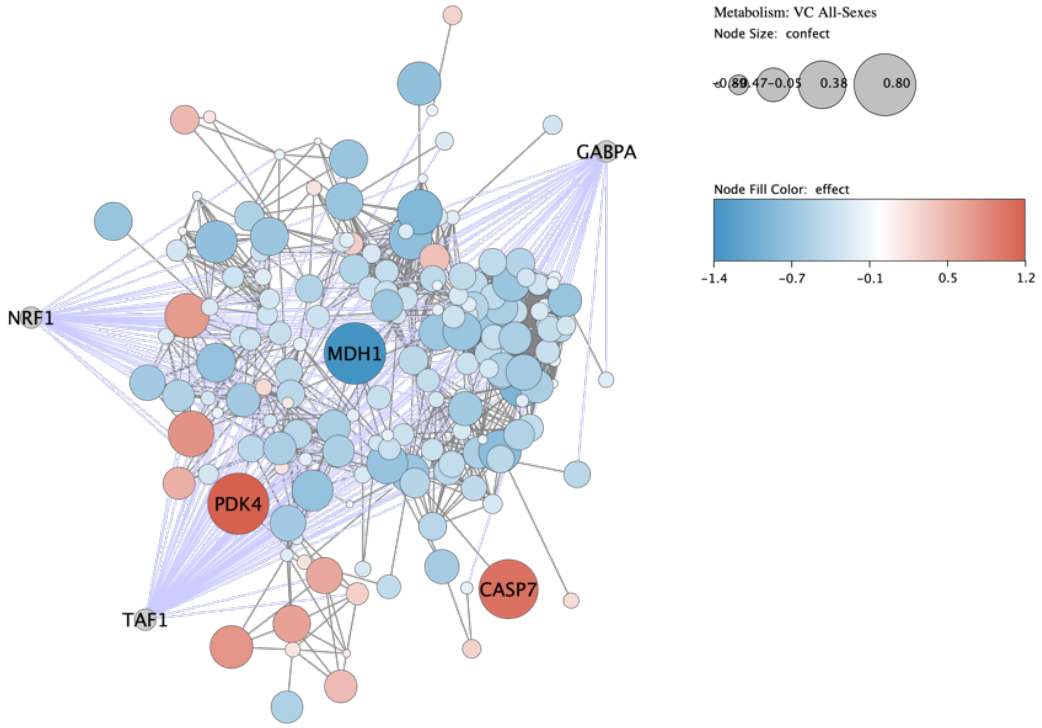


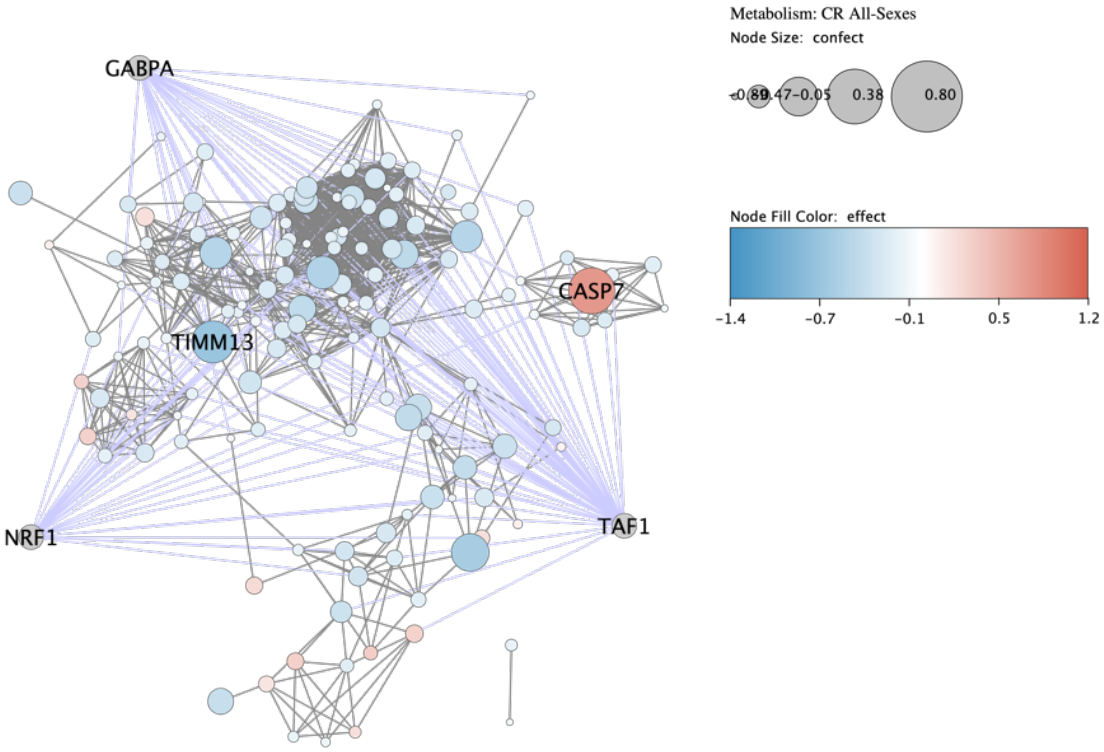
Figure 15. Metabolism Endotype PPI Network. A) PFC, B) MTG, C) VC, D) CR. Node size indicates differential expression confect score. Node color indicates differential expression effect. Grey nodes indicate TF target genes absent in the network. Purple lines represent TF-gene edges.

Figure 15 continued.

C



D



CONCLUSION

In this work, a pipeline was created to analyze microarray data and characterize differences between four brain regions and sexes in LOAD. In this process, enrichment analyses were used to identify functional pathways distinctive of AD: Cell Cycle G_1-S Phase, Inflammation, Dedifferentiation, Metabolism and Synaptic Signaling. Interestingly, the All-Sexes group was enriched in all brain regions for inflammation, dedifferentiation, cell cycle and metabolism, and synaptic signaling. However, the PFC and VC showed greater statistically significant enrichment of synaptic signaling, over the CR and VC.

When using the CERNO test, all brain regions resulted in similar functional enrichment, with results related to inflammation, dedifferentiation, cell cycle, metabolism, synaptic signaling and neurogenesis. Similar results were observed for the Females and Males. Slight preferential enrichment of inflammation, synaptic signaling and neurogenesis in the PFC; of synaptic signaling and neurogenesis in the MTG; of inflammation, synaptic signaling, neurogenesis in the VC, and of inflammation in the Cerebellum.

When using the fGSEA test, all brain regions showed increased inflammatory response, reduced oxidative phosphorylation, decreased synaptic signaling, increased differentiation and blood vessel morphogenesis. The PFC exhibited slight preferential enrichment of synaptic signaling, and of metabolism in the MTG and VC. Females indicated similar trends of functional enrichment: increased blood vessel development and inflammation. The PFC and MTG in the female group also exhibited slight preferential negative enrichment of synaptic signaling. For the males, a similar trend was observed: blood vessel development and inflammation were activated across all brain regions. Synaptic signaling was suppressed in the PFC and MTG, and metabolism was suppressed in the MTG and VC.

Additionally, TFs were identified, specific to each brain region and endotype, then used to generate PPI networks to allow visualization of the gene protein-protein and TF-gene interactions. These networks reinforced the results observed previously: the PFC and VC are more greatly affected by LOAD-associated disease endotypes. Interestingly, the MTG appears to have the most significant enrichment related to the suppression of metabolism, whereas the CR appears to have the most significant enrichment related to the suppression of metabolism and activation of inflammation, although though both the MTG and the CR are comparatively less affected than either the PFC or VC.

FUTURE DIRECTIONS

The future of this work has many potential avenues. The initial next steps may involve exploring additional endotypes. While the five represented in this work were highly represented across the results, there were many other endotypes that would also be of interest to explore, such as Cell Cycle G_2-M Phase and Neurogenesis. There are additionally many other TFs that may be incorporated into the PPIs to further explore the interaction networks across brain regions and sexes. Additionally, subnetworks may be created and analyzed from the generated PPI networks, to explore specific connections in more detail. Regardless of the direction taken, there are many possibilities left to unearth to allow for better understanding of LOAD mechanisms.

REFERENCES

Agod, Z., Pazmandi, K., Bencze, D., Vereb, G., Biro, T., Szabo, A., Rajnavolgyi, E., Bacsi, A., Engel, P., & Lanyi, A. (2018). Signaling lymphocyte activation molecule family 5 enhances autophagy and fine-tunes cytokine response in monocyte-derived dendritic cells via stabilization of interferon regulatory factor 8. *Frontiers in immunology*, *9*, 62.

Alvarez, M. (2020). *Virtual Inference of Protein Activity by Enriched Regulon Analysis.* Retrieved 2020, from Bioconductor.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., & Eppig, J. T. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25-29.

Bettelli, E., Dastrange, M., & Oukka, M. (2005). Foxp3 interacts with nuclear factor of activated T cells and NF-κB to repress cytokine gene expression and effector functions of T helper cells. *Proceedings of the National Academy of Sciences*, *102*(14), 5138-5143.

Cartwright, P., MuÈller, H., Wagener, C., Holm, K., & Helin, K. (1998). E2F-6: a novel member of the E2F family is an inhibitor of E2F-dependent transcription. *Oncogene*, *17*(5), 611-623.

Chai, J., Wu, Q., Shiozaki, E., Srinivasula, S. M., Alnemri, E. S., & Shi, Y. (2001). Crystal structure of a procaspase-7 zymogen: mechanisms of activation and substrate binding. *Cell*, *107*(3), 399-407.

Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, *14*(1), 128.

Chen, K., Liu, J., Liu, S., Xia, M., Zhang, X., Han, D., Jiang, Y., Wang, C., & Cao, X. (2017). Methyltransferase SETD2-mediated methylation of STAT1 is critical for interferon antiviral activity. *Cell*, *170*(3), 492-506. e414.

Chèneby, J., Ménétrier, Z., Mestdagh, M., Rosnet, T., Douida, A., Rhalloussi, W., Bergon, A., Lopez, F., & Ballester, B. (2020). ReMap 2020: A database of regulatory regions from an integrative analysis of Human and Arabidopsis DNA-binding sequencing experiments. *Nucleic acids research*, *48*(D1), D180-D188.

Consortium, G. O. (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic acids research*, *47*(D1), D330-D338.

CRH corticotropin releasing hormone. (2020, November 24). Retrieved from NCBI NLM NIH: https://www.ncbi.nlm.nih.gov/gene/1392#gene-expression

Cyclin-dependent kinase 2. (2020, November 23). Retrieved from PubChem: https://pubchem.ncbi.nlm.nih.gov/protein/P24941

Davis, C. A., Hitz, B. C., Sloan, C. A., Chan, E. T., Davidson, J. M., Gabdank, I., Hilton, J. A., Jain, K., Baymuradov, U. K., & Narayanan, A. K. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic acids research*, *46*(D1), D794-D801.

Desdouets, C., Matesic, G., Molina, C. A., Foulkes, N. S., Sassone-Corsi, P., Brechot, C., & Sobczak-Thepot, J. (1995). Cell cycle regulation of cyclin A gene expression by the cyclic AMP-responsive transcription factors CREB and CREM. *Molecular and cellular biology*, *15*(6), 3301-3309.

Fan, Z., Beresford, P. J., Oh, D. Y., Zhang, D., & Lieberman, J. (2003). Tumor suppressor NM23-H1 is a granzyme A-activated DNase during CTL-mediated apoptosis, and the nucleosome assembly protein SET is its inhibitor. *Cell*, *112*(5), 659-672.

Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D., & Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome research*, *29*(8), 1363-1375.

García-Gutiérrez, L., Delgado, M. D., & León, J. (2019). MYC oncogene contributions to release of cell cycle brakes. *Genes*, *10*(3), 244.

Gopalakrishnan, L., & Scarpulla, R. C. (1995). Structure, expression, and chromosomal assignment of the human gene encoding nuclear respiratory factor 1. *Journal of Biological Chemistry*,*270*(30), 18019-18025.

Harrison, P. (2020, October 27). *An overview of topconfects.* Retrieved 2020, from Bioconductor.

Harrison, P. (2020, November 19). *Top Confident Effect Sizes.* Retrieved 2020, from Bioconductor.

Harrison, P. F., Pattison, A. D., Powell, D. R., & Beilharz, T. H. (2019). Topconfects: a package for confident effect sizes in differential expression analysis provides a more biologically useful ranked gene list. *Genome biology*, *20*(1), 1-12.

Hotta, K., Hosaka, M., Tanabe, A., & Takeuchi, T. (2009). Secretogranin II binds to secretogranin III and forms secretory granules with orexin, neuropeptide Y, and POMC. *Journal of endocrinology*, *202*(1), 111.

Korotkevich, G., Sukhov, V., & Sergushichev, A. (2019). Fast gene set enrichment analysis. *BioRxiv*, 060012.

Korotkevich, G., Sukhov, V., Budin, N., & Sergushichev, A. (2020, November 20). *Fast Gene Set Enrichment Analysis.* Retrieved 2020, from Bioconductor.

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S. L., Jagodnik, K. M., & Lachmann, A. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, *44*(W1), W90-W97.

Lachmann, A., Xu, H., Krishnan, J., Berger, S. I., Mazloom, A. R., & Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. *Bioinformatics*, *26*(19), 2438-2444.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. *Cell systems*, *1*(6), 417-425.

Lunyak, V. V., Burgess, R., Prefontaine, G. G., Nelson, C., Sze, S.-H., Chenoweth, J., Schwartz, P., Pevzner, P. A., Glass, C., & Mandel, G. (2002). Corepressor-dependent silencing of chromosomal regions encoding neuronal genes. *Science*, *298*(5599), 1747-1752.

Luque, R. M., & Kineman, R. D. (2018). Neuronostatin exerts actions on pituitary that are unique from its sibling peptide somatostatin. *Journal of Endocrinology*, *237*(3), 217-227.

Ma, L., Huang, C., Wang, X.-J., Xin, D. E., Wang, L.-s., Zou, Q. C., Ya-nan, S. Z., Tan, M.-d., Wang, Y.-m., & Zhao, T. C. (2017). Lysyl oxidase 3 is a dual-specificity enzyme involved in STAT3 deacetylation and deacetylimination modulation. *Molecular cell*,*65*(2), 296-309.

NCBI. (2020, November 24). *MDH1 malate dehydrogenase 1*. Retrieved from NCBI NLM NIH: https://www.ncbi.nlm.nih.gov/gene/4190

Negoro, H., Kanematsu, A., Doi, M., Suadicani, S. O., Matsuo, M., Imamura, M., Okinami, T., Nishikawa, N., Oura, T., & Matsui, S. (2012). Involvement of urinary bladder Connexin43 and the circadian clock in coordination of diurnal micturition rhythm. *Nature communications*, *3*(1), 1-10.

Piras, I. S., Krate, J., Delvaux, E., Nolz, J., Mastroeni, D. F., Persico, A. M., Jepsen, W. M., Beach, T. G., Huentelman, M. J., & Coleman, P. D. (2019). Transcriptome changes in the Alzheimer's disease middle temporal gyrus: importance of RNA metabolism and mitochondria-associated membrane genes. *Journal of Alzheimer's Disease*, *70*(3), 691-713.

Reeves, S. A., Helman, L. J., Allison, A., & Israel, M. A. (1989). Molecular cloning and primary structure of human glial fibrillary acidic protein. *Proceedings of the National Academy of Sciences*,*86*(13), 5178-5182.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, *43*(7), e47-e47.

Ryckman, C., Vandal, K., Rouleau, P., Talbot, M., & Tessier, P. A. (2003). Proinflammatory activities of S100: proteins S100A8, S100A9, and S100A8/A9 induce neutrophil chemotaxis and adhesion. *The Journal of Immunology*, *170*(6), 3233-3242.

Sells, M. A., Knaus, U. G., Bagrodia, S., Ambrose, D. M., Bokoch, G. M., & Chernoff, J. (1997). Human p21-activated kinase (Pak1) regulates actin organization in mammalian cells. *Current Biology*, *7*(3), 202-210.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), 2498-2504.

Smyth, G., Hu, Y., Ritchie, M., Silver, J., Wettenhall, J., McCarthy, D., Wu, D., Shi, W., Phipson, B., Lun, A., Thorne, N., Oshlack, A., de Graaf, C., Chen, Y., Langaas, M., Ferkingstad, E., Davy, M., Pepin, F., & Choi, D. (2020, October 19). *Linear Models for Microarray Data.* Retrieved 2020, from Bioconductor.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., & Lander, E. S. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, *102*(43), 15545-15550.

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., & Bork, P. (2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic acids research*,*47*(D1), D607-D613.

Wang, S.-C., Oelze, B., & Schumacher, A. (2008). Age-specific epigenetic drift in late-onset Alzheimer's disease. *PloS one*,*3*(7), e2698.

Weiner, J. (2020, October 2). *Feature Set Enrichment Analysis for Metabolomics and Transcriptomics.* Retrieved 2020, from CRAN.

Wellner, U., Schubert, J., Burk, U. C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., & Zur Hausen, A. (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nature cell biology*, *11*(12), 1487-1495.

Wu, D., & Smyth, G. K. (2012). Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic acids research*,*40*(17), e133-e133.

Yang, Z.-F., Drumea, K., Mott, S., Wang, J., & Rosmarin, A. G. (2014). GABP transcription factor (nuclear respiratory factor 2) is required for mitochondrial biogenesis. *Molecular and cellular biology*, *34*(17), 3194-3201.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A. A., Zhang, C., Xie, T., Tran, L., & Dobrin, R. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, *153*(3), 707-720.

Zhao, D. M. (2020). *dbemt(v1)*. Retrieved from Epithelial-Mesnchymal Transition Gene Database.

Zhao, M., Kong, L., Liu, Y., & Qu, H. (2015, June 2015). dbEMT: a literature-based resource for Epithelial-Mesenchymal Transition genes. *Scientific Reports, 23*(5).

Zyla, J., Marczyk, M., Domaszewska, T., Kaufmann, S. H., Polanska, J., & Weiner 3rd, J. (2019). Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. *Bioinformatics*, *35*(24), 5146-5154.

APPENDICES

Appendix A. Differential Expression Results: *topconfects* vs *eBayes*

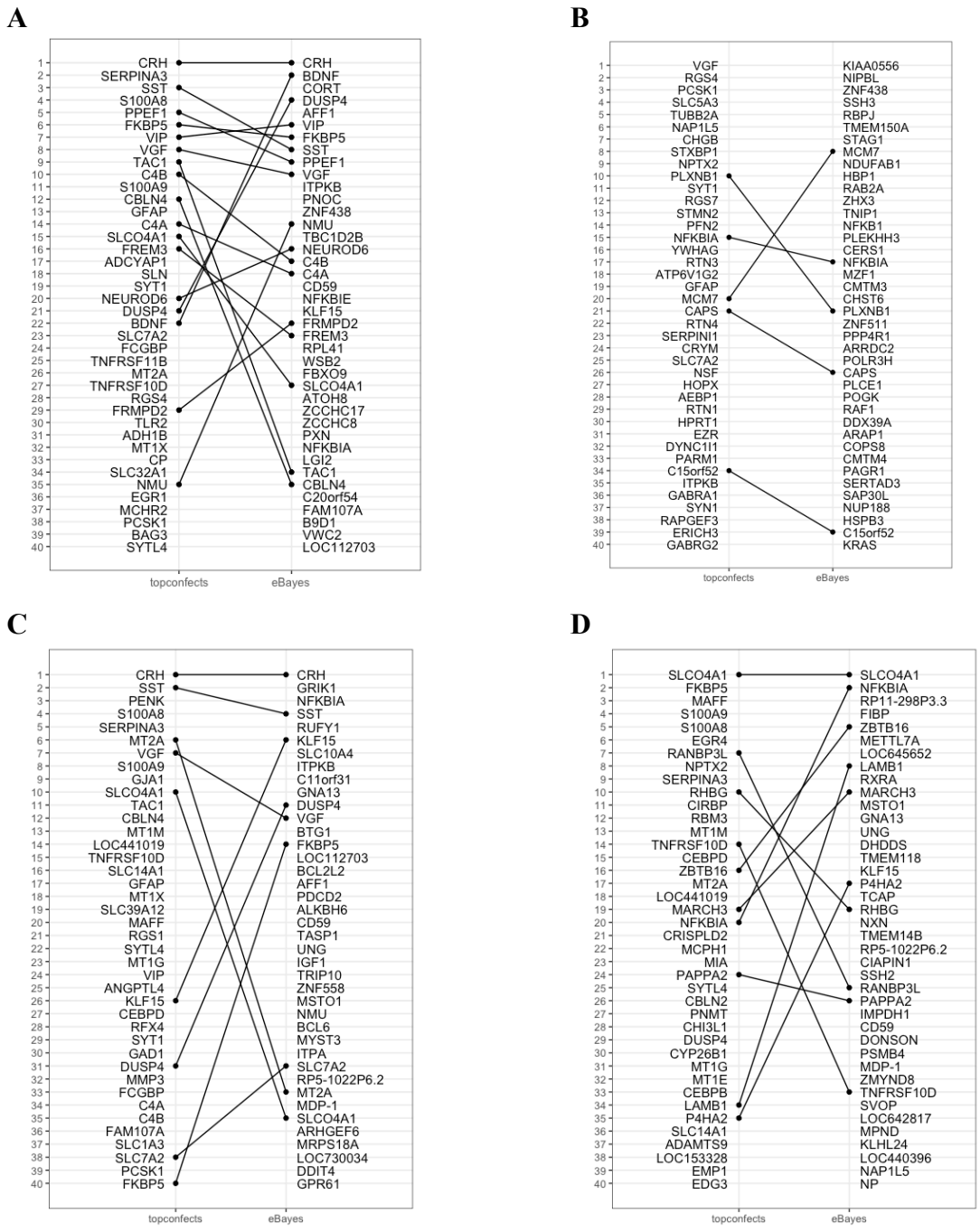


Figure 16. Rank plot comparing top 40 topconfects and eBayes DEGs for All Sexes group. A) PFC, B) MTG, C) VC, D) CR.

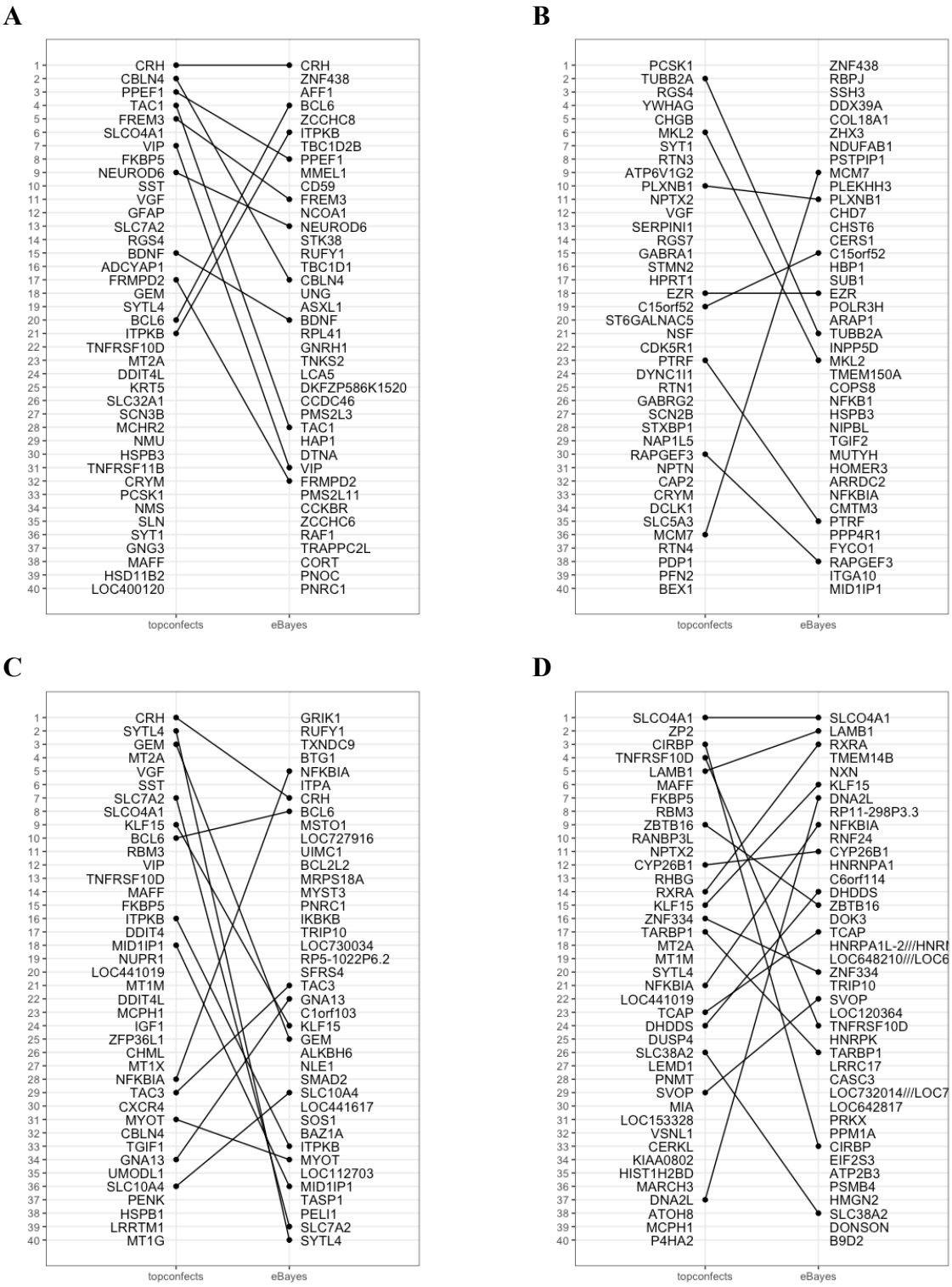


Figure 17. Rank plot comparing top 40 topconfects and eBayes DEGs for Female group. A) PFC, B) MTG, C) VC, D) CR.

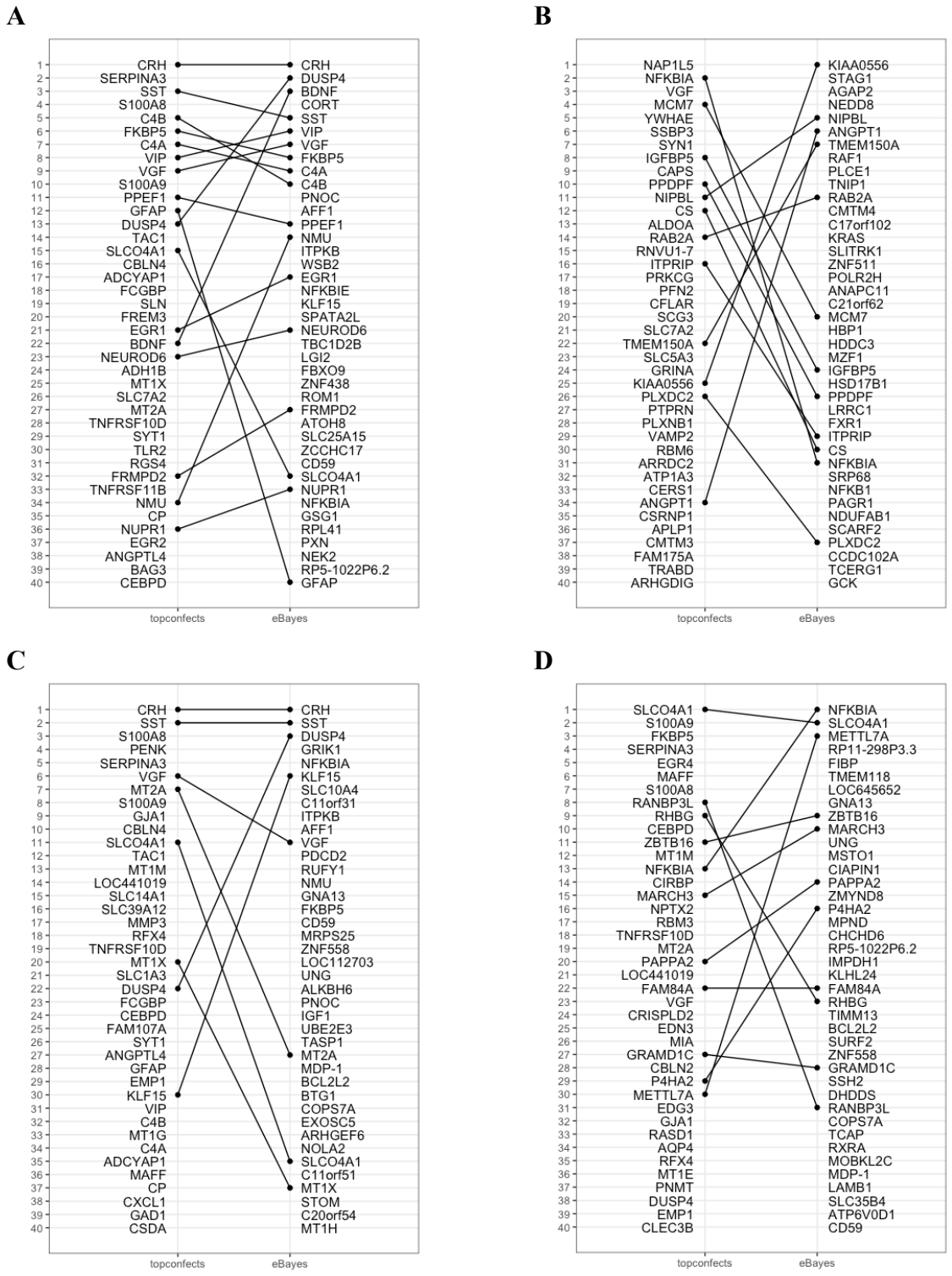


Figure 18. Rank plot comparing top 40 topconfects and eBayes DEGs for Male group. A) PFC, B) MTG, C) VC, D) CR.

Appendix B. Enrichment Test Results: CERNO

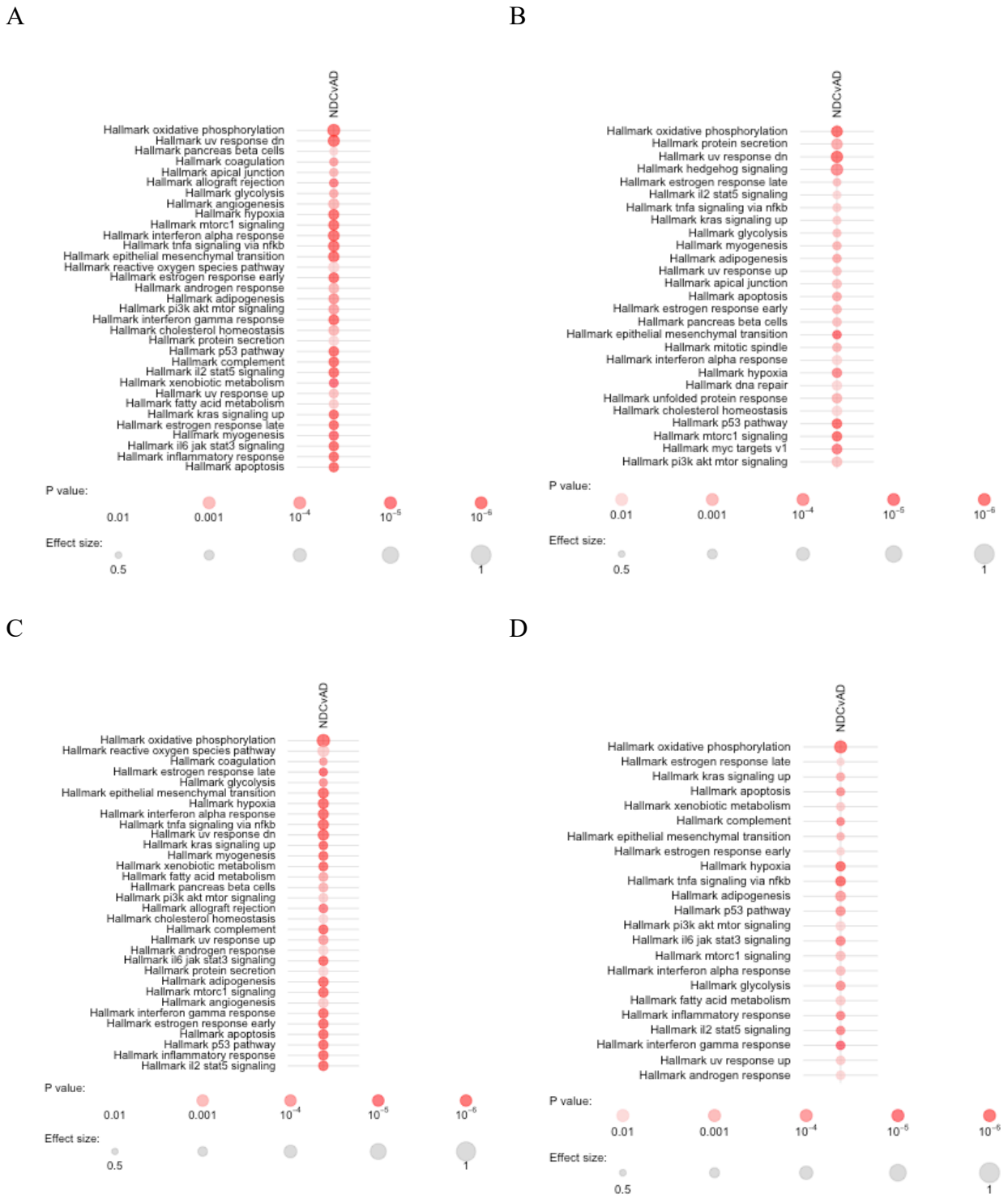
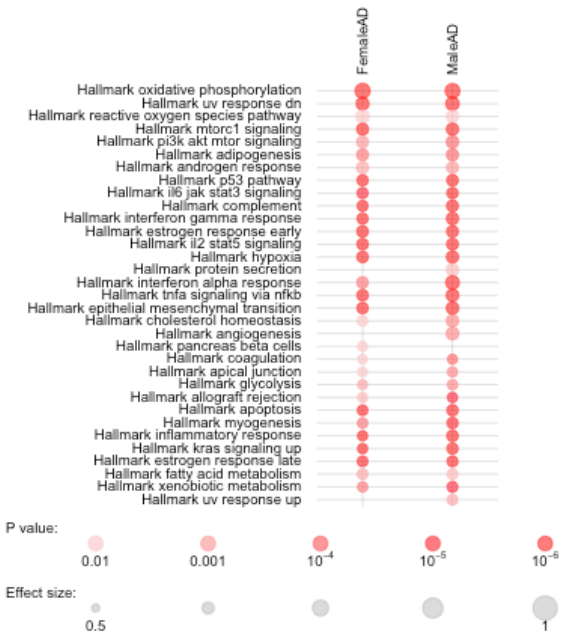
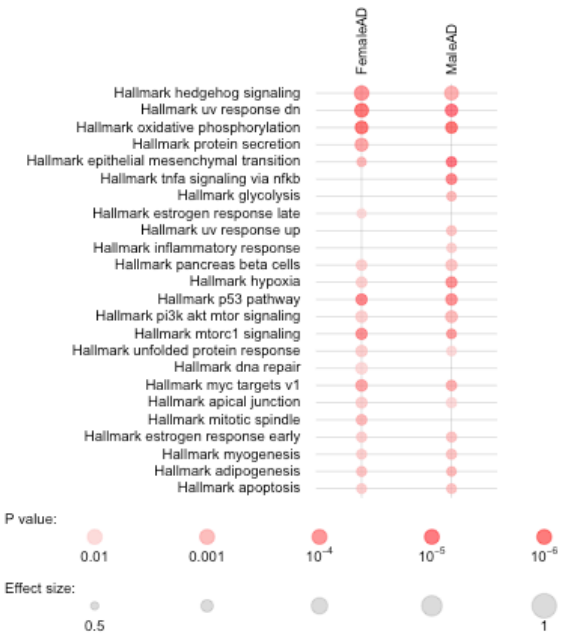


Figure 19. Panel plots of CERNO + Hallmark results for the All-Sexes group. A) PFC, B) MTG, C) VC, D) CR.

A



B



C

D

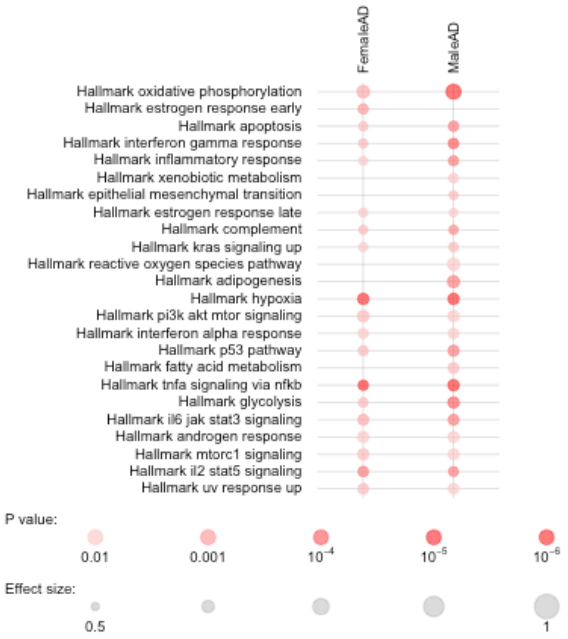
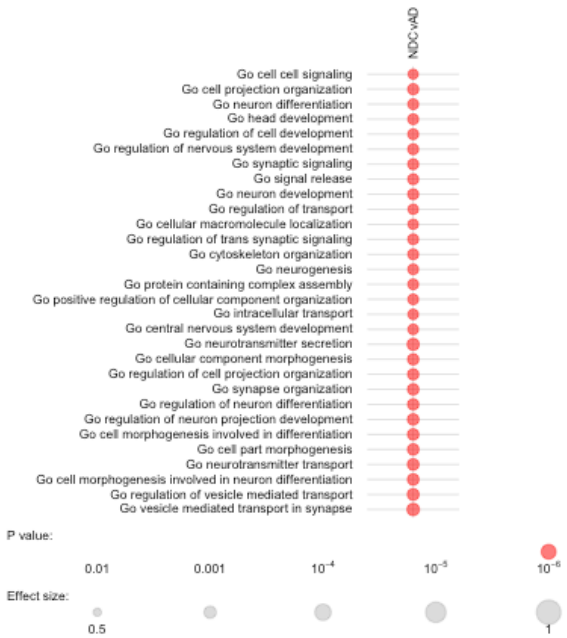


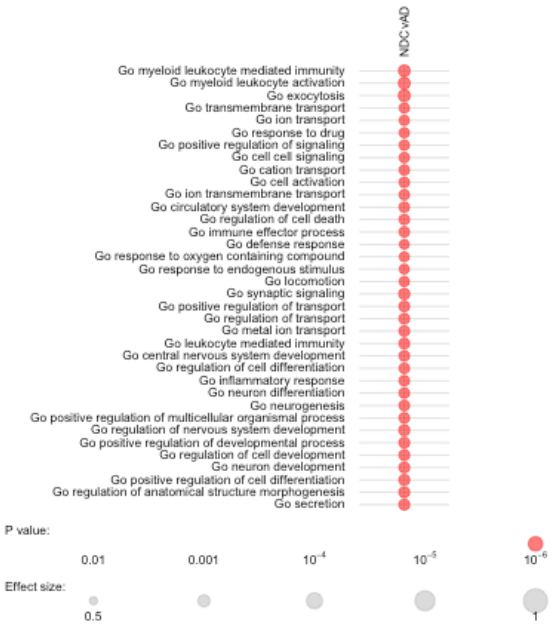
Figure 20. Panel plot comparing CERNO + Hallmark results for Female and Male groups. A) PFC, B) MTG, C) VC, D) CR.

A

B



C



D

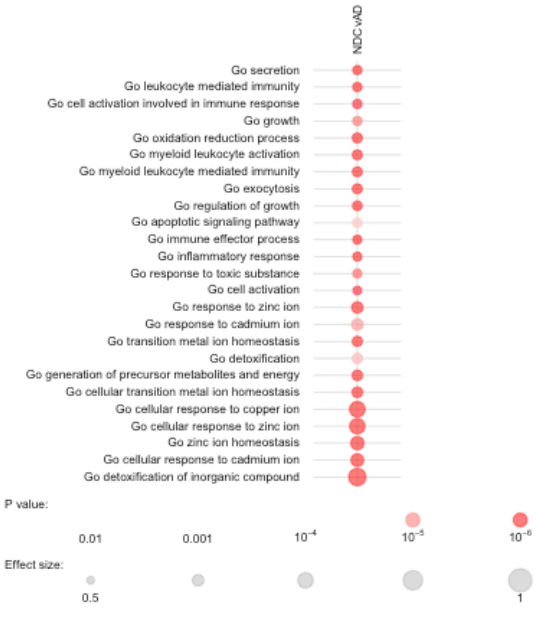


Figure 21. Panel of CERNO + GOBP results for All-Sexes group. A) PFC, B) MTG, C) VC, D) CR.

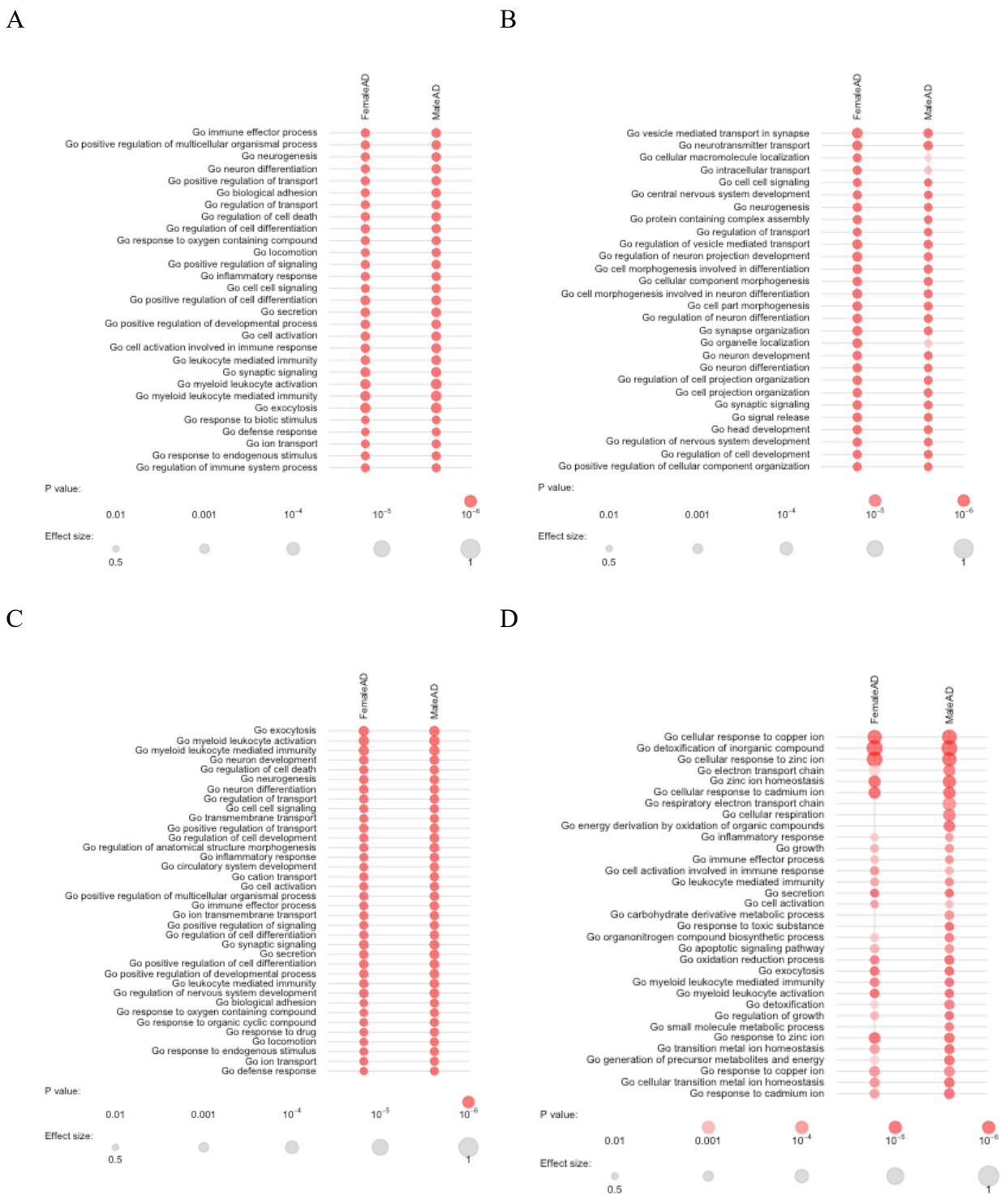
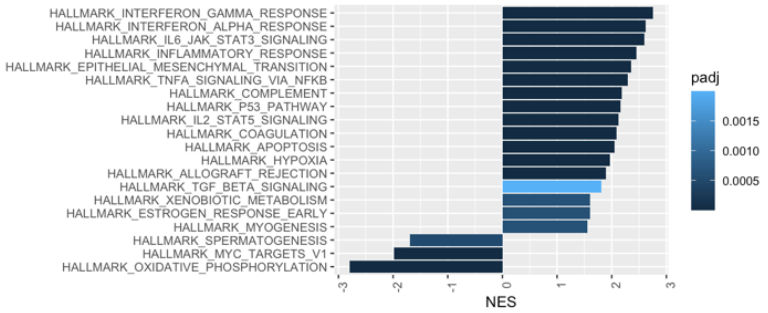


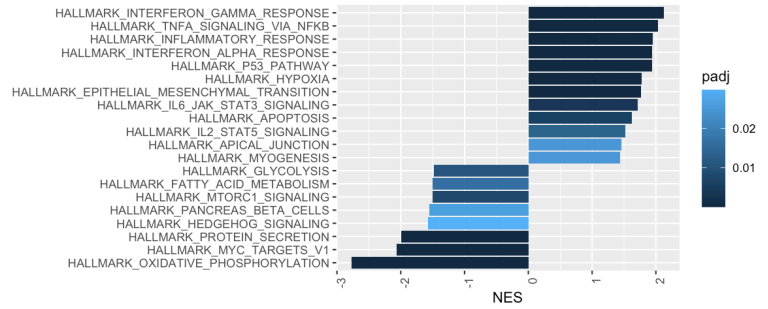
Figure 22. Panel plot comparing CERNO + GOBP results for Female and Male groups. A) PFC, B) MTG, C) VC, D) CR.

Appendix C. Enrichment Test Results: fGSEA + Hallmark

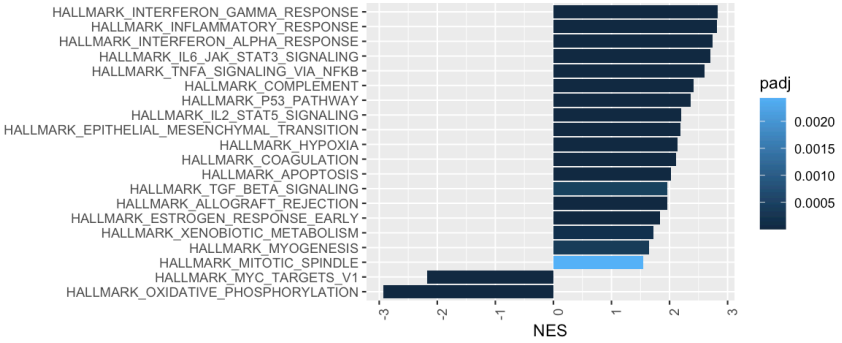
A



B



C



D

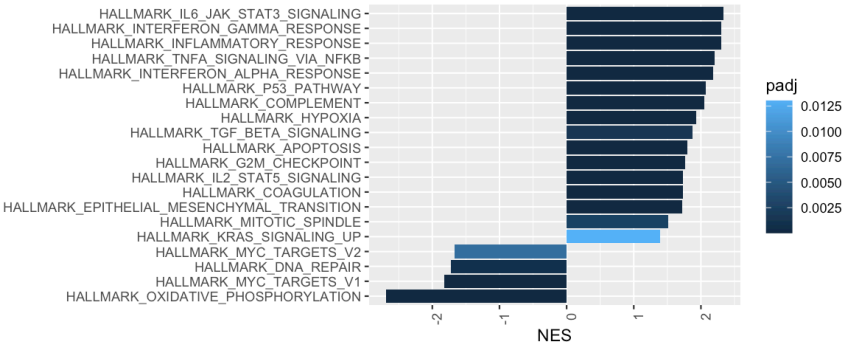
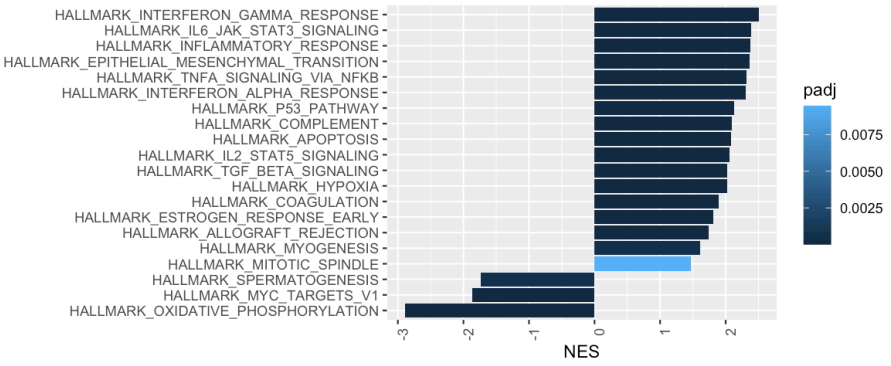
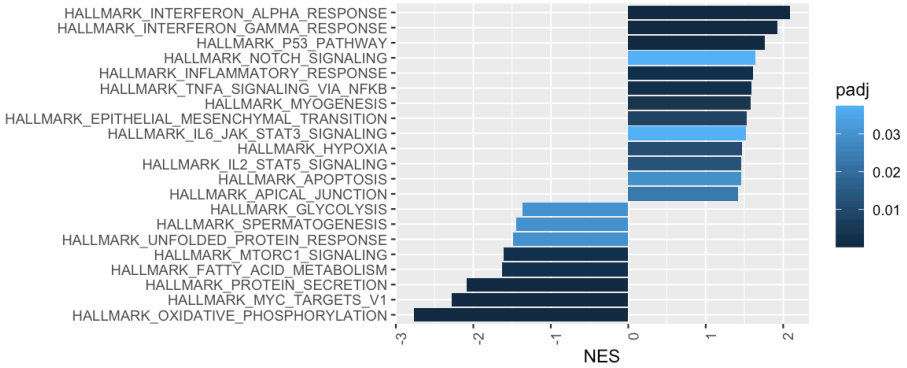


Figure 23. Top fGSEA + Hallmark results for the All-Sexes group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

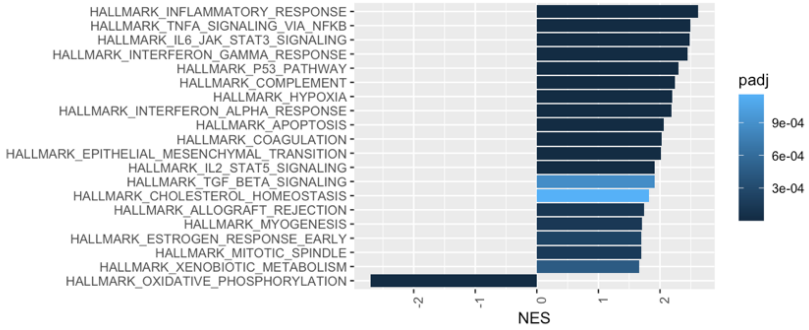
A



B



C



D

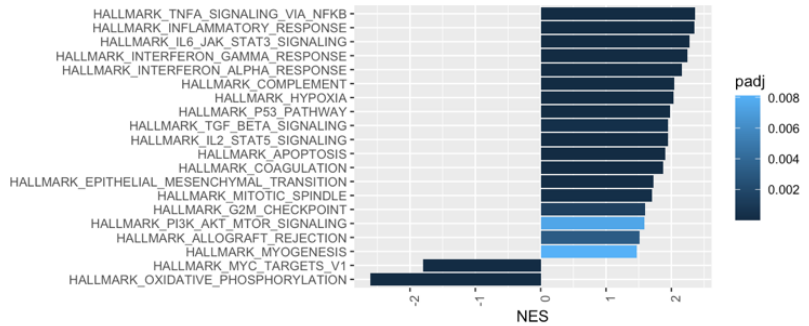
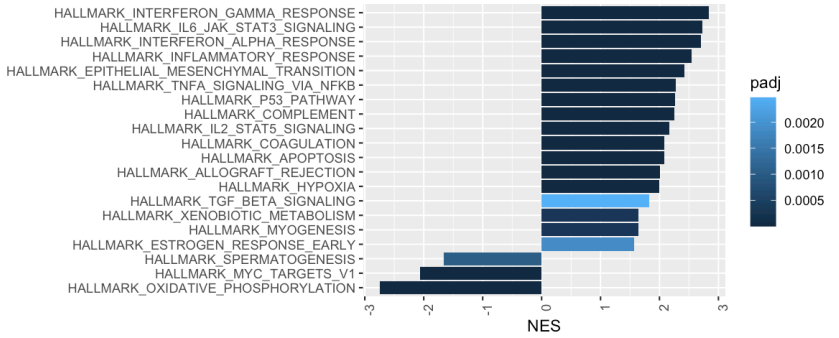
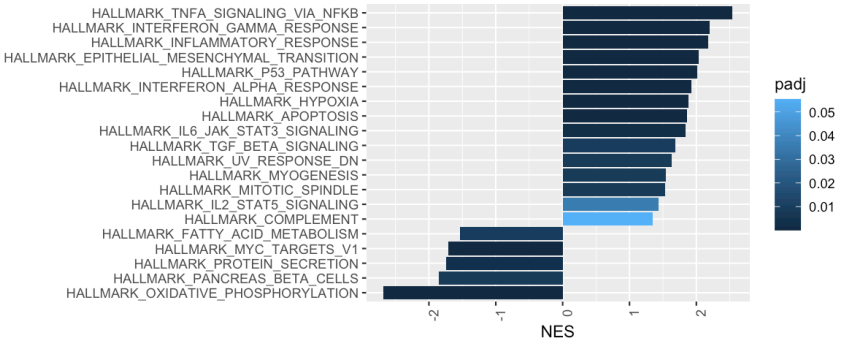


Figure 24. Top fGSEA + Hallmark results for the Female group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

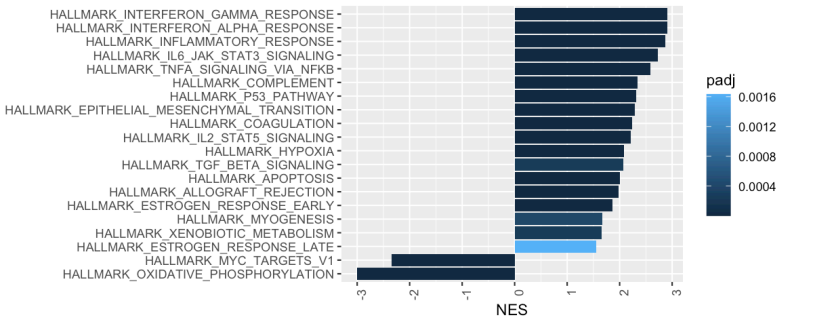
A



B



C



D

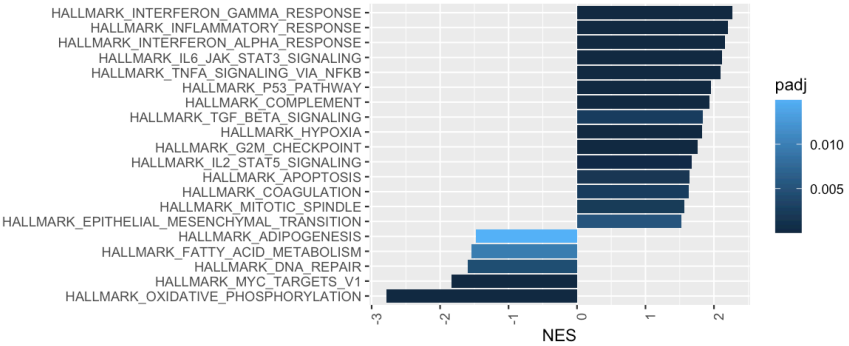
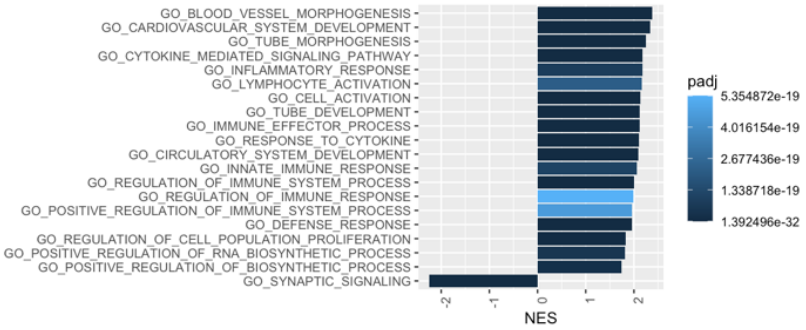


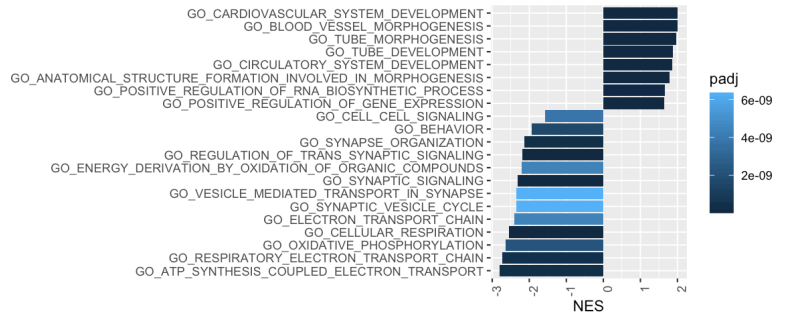
Figure 25. Top fGSEA + Hallmark results for the Male group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

Appendix D. Enrichment Test Results: fGSEA + GOBP

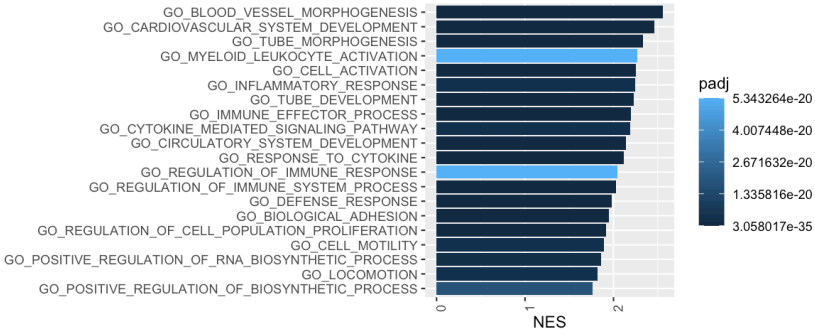
A



B



C



D

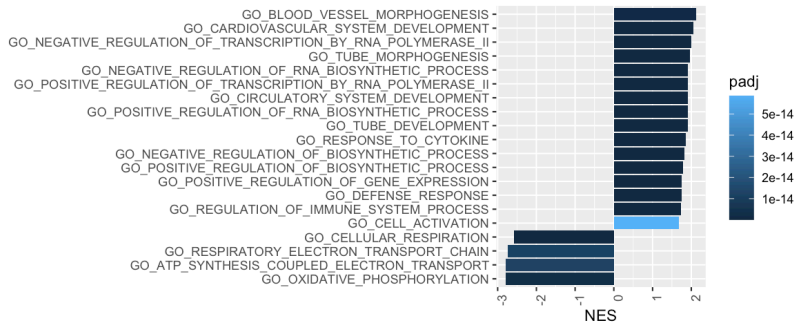
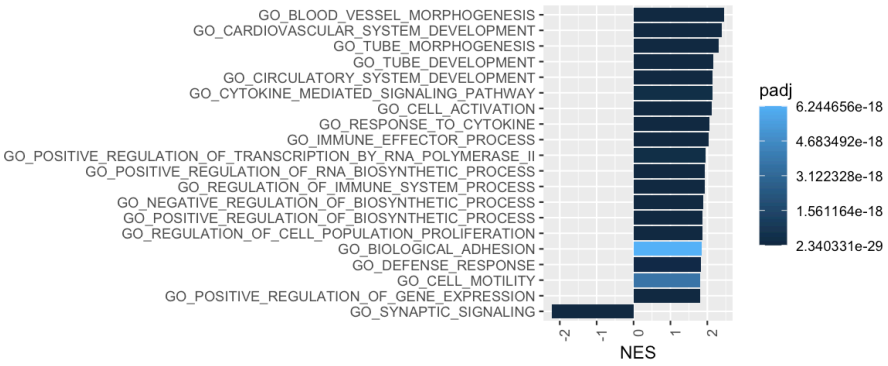
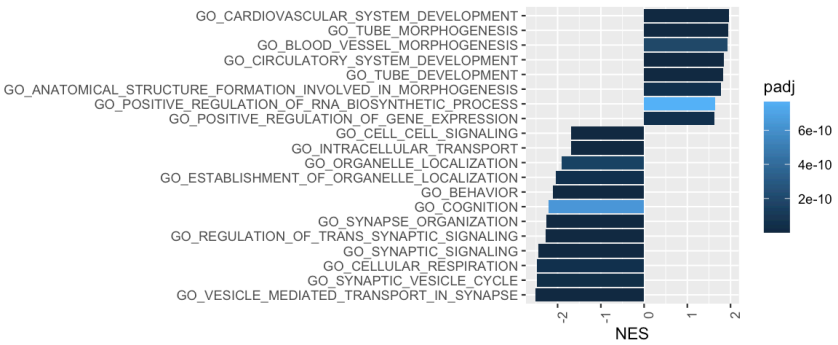


Figure 26. Top fGSEA + GOBP results for the All-Sexes group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

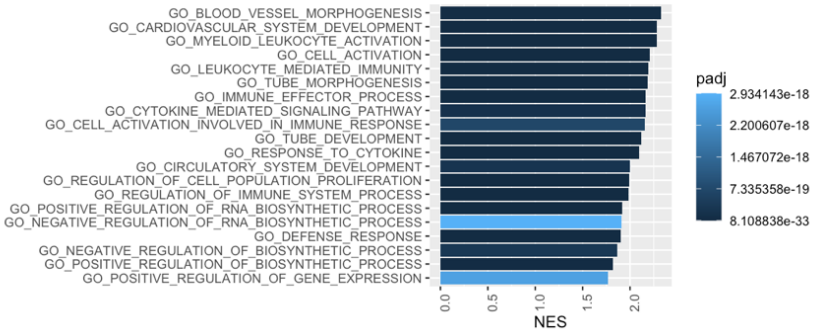
A



B



C



D

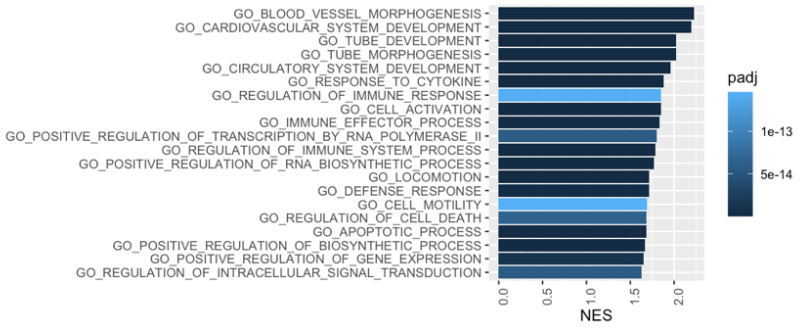
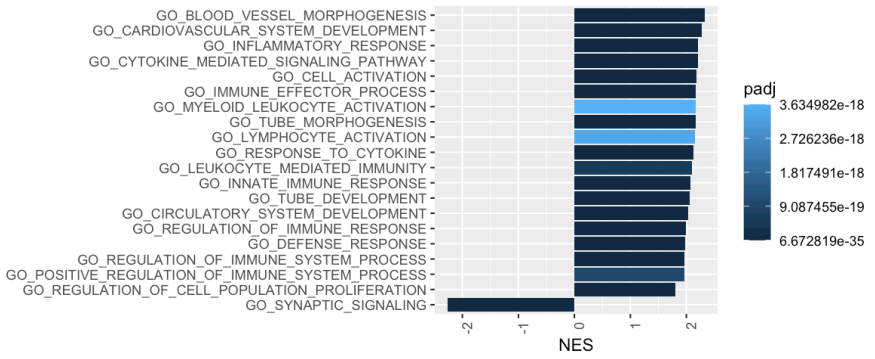
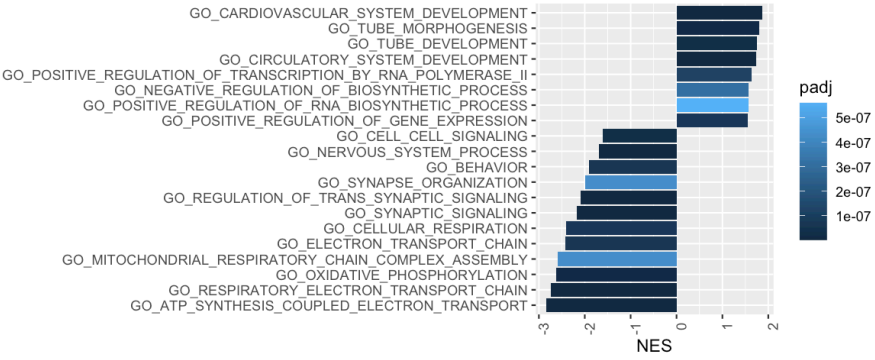


Figure 27. Top fGSEA + GOBP results for the Female group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

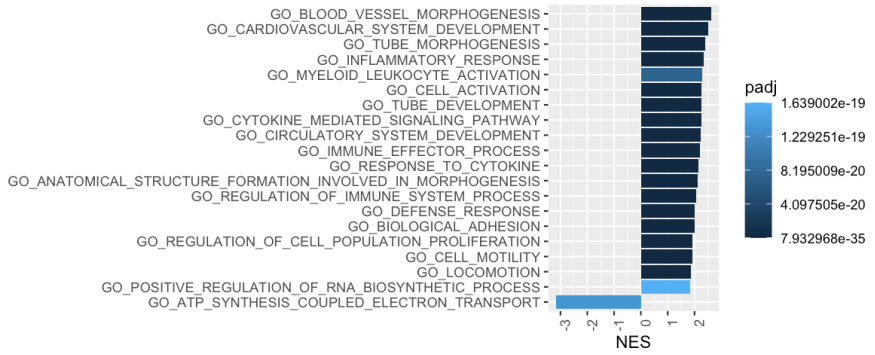
A



B



C



D

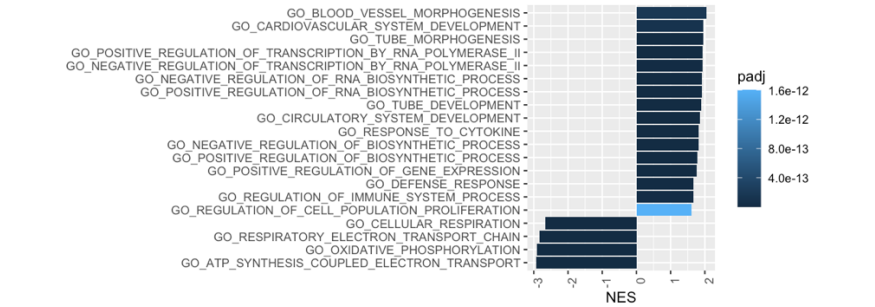
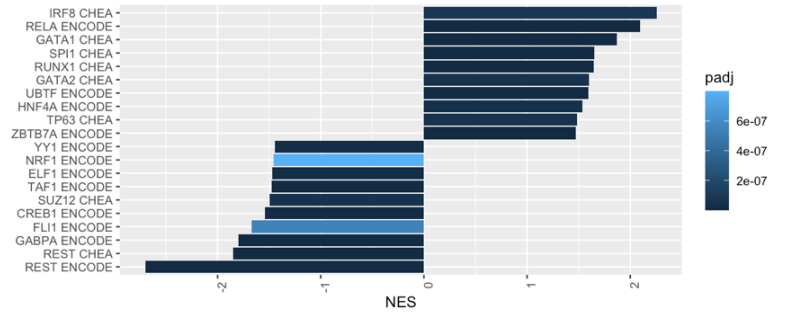


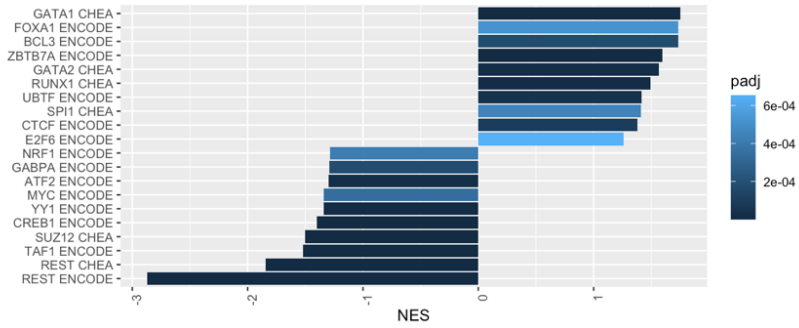
Figure 28. Top fGSEA + GOBP results for the Male group. Bar plots summarize the twenty pathways with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

Appendix E. TF Enrichment Test Results: fGSEA + ECC

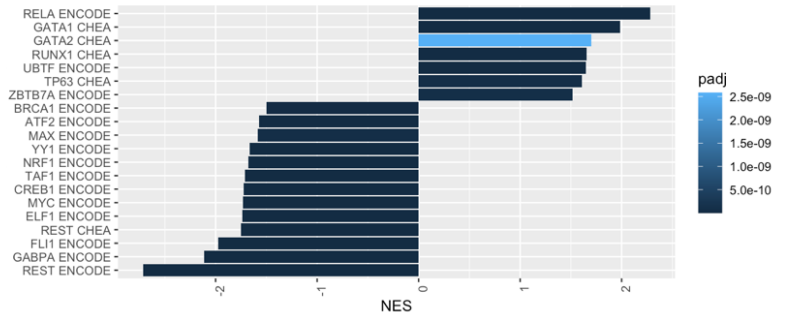
A



B



C



D

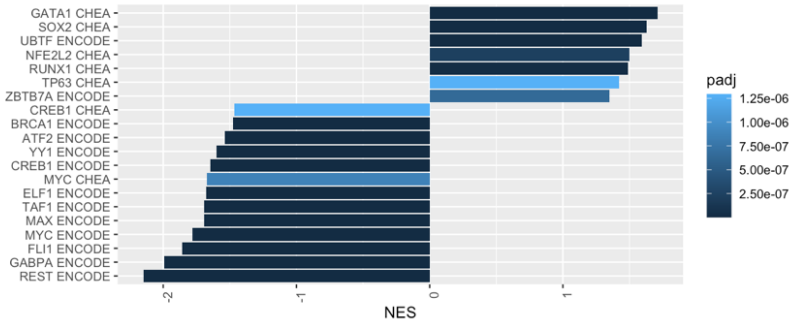
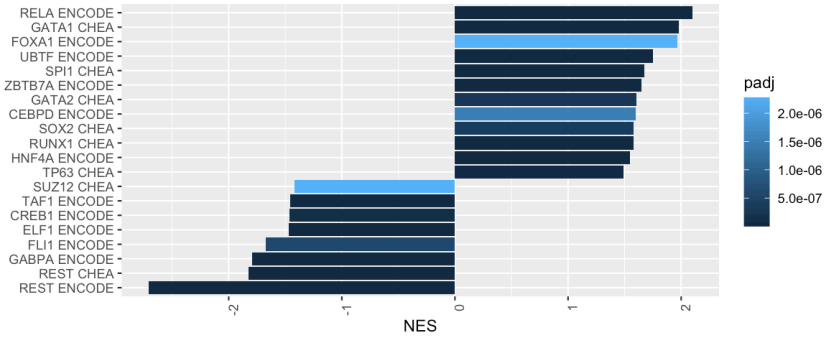
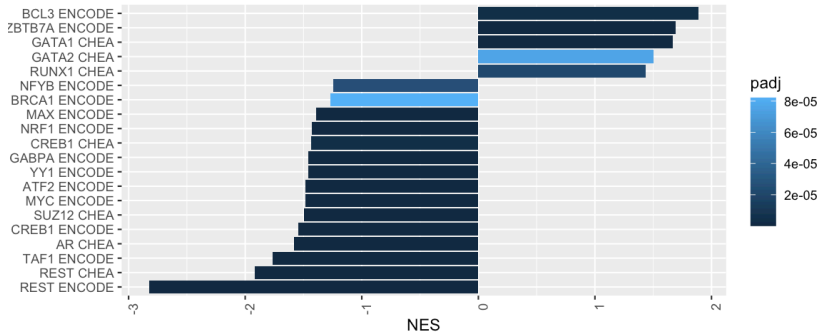


Figure 29. Top TFs of fGSEA + ECC results for All Sexes group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

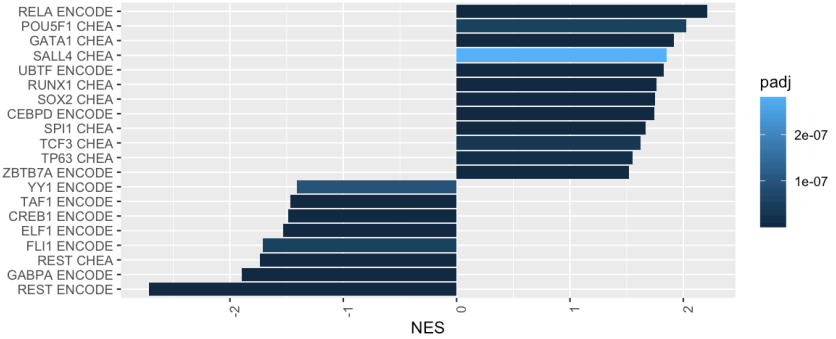
A



B



C



D

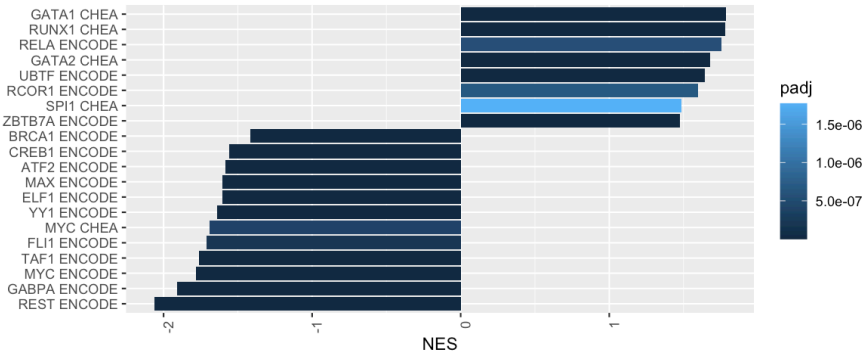
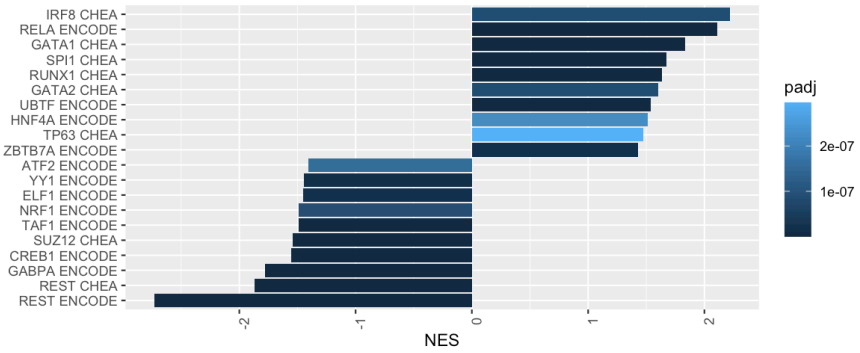
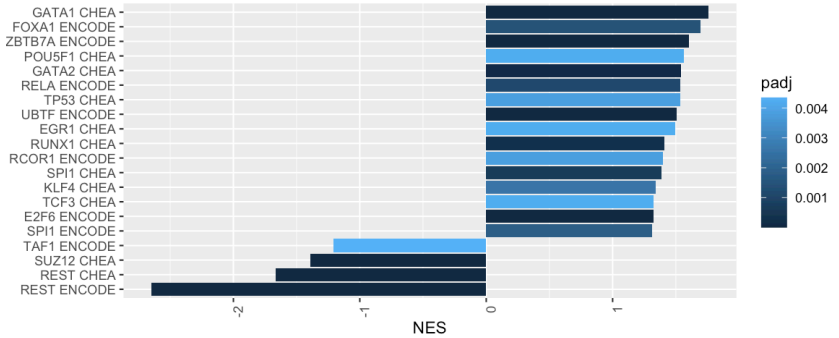


Figure 30. Top TFs of fGSEA + ECC results for the Female group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

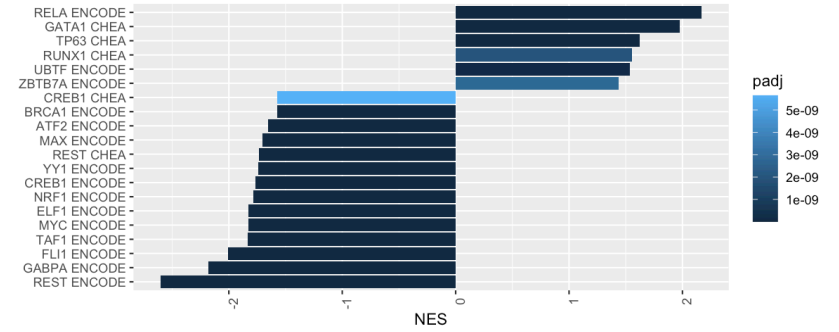
A



B



C



D

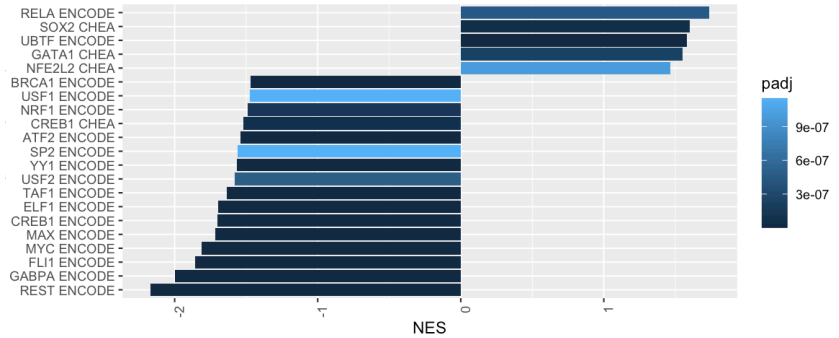
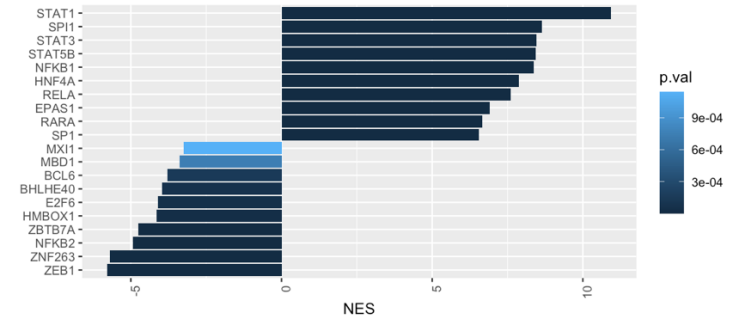


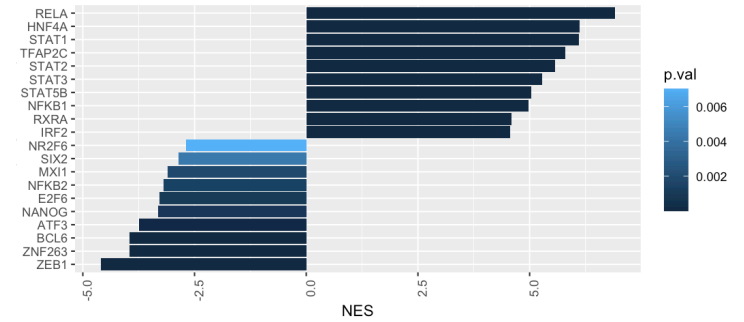
Figure 31. Top TFs of fGSEA + ECC results for the Male group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

Appendix F. TF Enrichment Test Results: DoRothEA

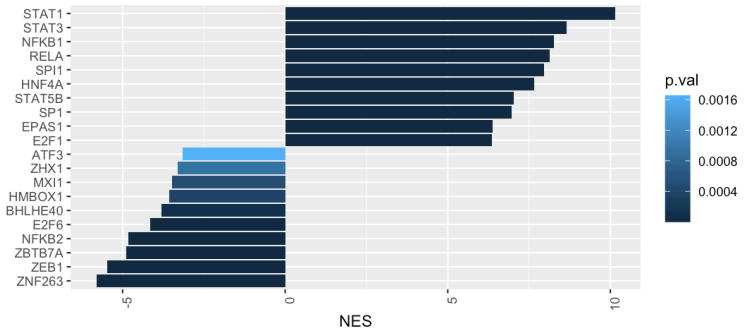
A



B



C



D

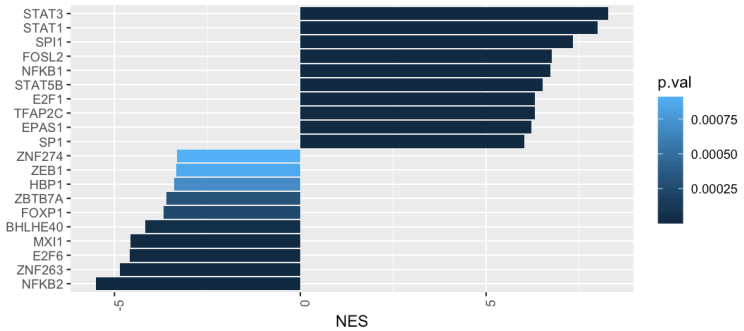
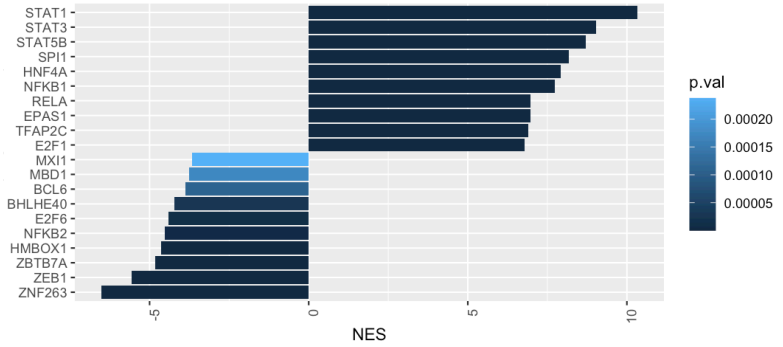
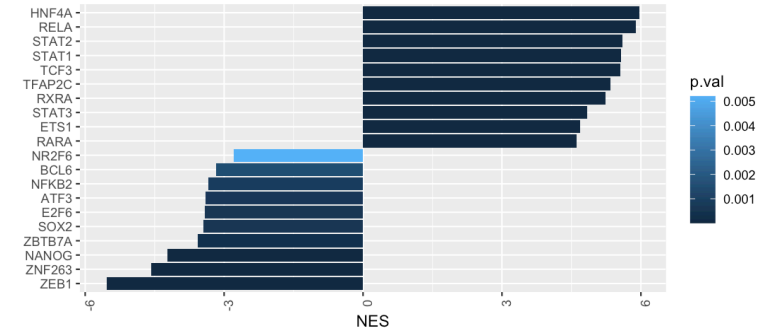


Figure 32. Top regulons of DoRothEA results for All Sexes group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

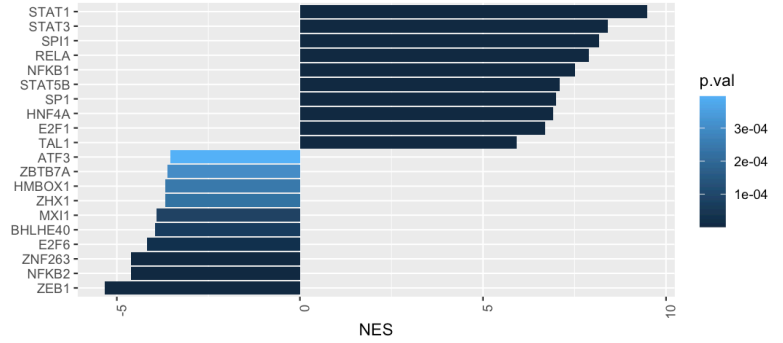
A



B



C



D

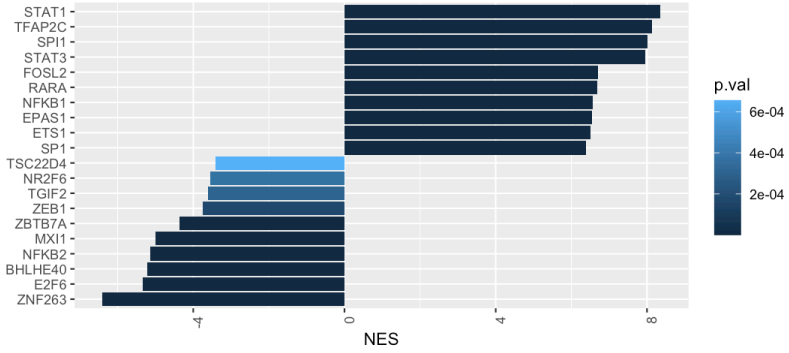
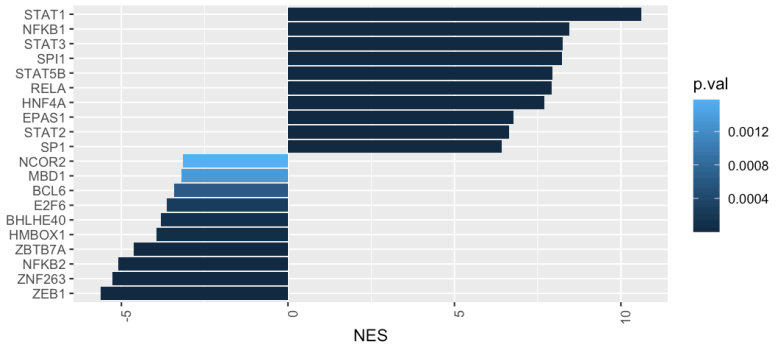
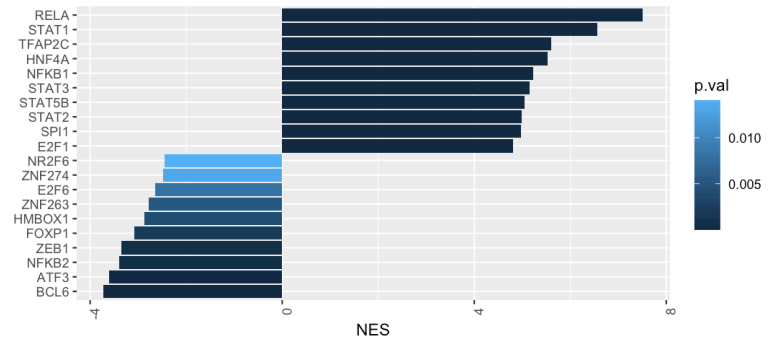


Figure 33. Top regulons of DoRothEA results for Female group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR

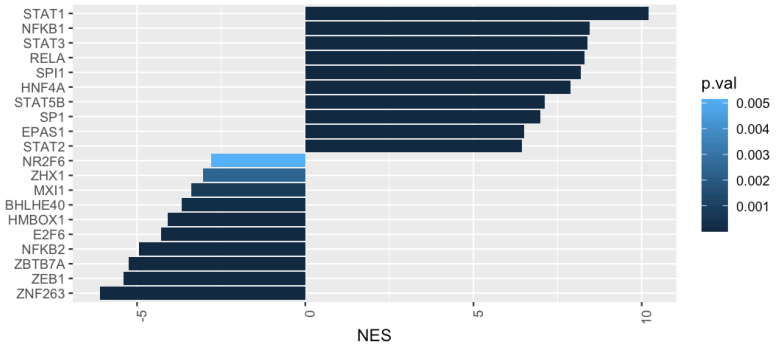
A



B



C



D

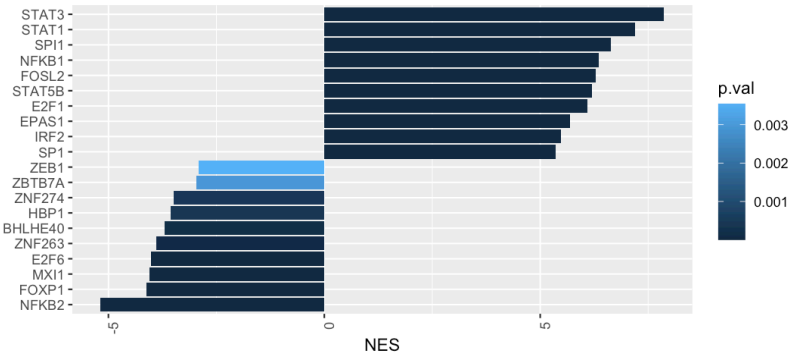


Figure 34. Top regulons of DoRothEA results for Male group. Bar plots summarize the twenty regulons with the lowest p-values, ordered by NES. Bar color indicates p-value. A) PFC, B) MTG, C) VC, D) CR