COMPARATIVE ANALYSIS OF BRAIN CELL CULTURES AND TISSUES IN ALZHEIMER'S DISEASE BASED ON DIFFERENTIAL EXPRESSION AND GENE SET ENRICHMENT

A THESIS SUBMITTED TO

THE GRADUATE SCHOOL OF INFORMATICS OF

THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

HÜSEYİN CAHİT BURDUROĞLU

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

IN

THE DEPARTMENT OF BIONFORMATICS

JANUARY 2023

COMPARATIVE ANALYSIS OF BRAIN CELL CULTURES AND TISSUES IN ALZHEIMER'S DISEASE BASED ON DIFFERENTIAL EXPRESSION AND GENE SET ENRICHMENT

Submitted by HÜSEYİN CAHİT BURDUROĞLU in partial fulfillment of the requirements for the degree of **Master of Science in Health Informatics Department, Middle East Technical University** by,

Prof. Banu Günel Kılıç
Dean, **Graduate School of Informatics** _____

Assoc. Prof. Yeşim Aydın Son
Head of Department, **Health Informatics** _____

Assoc. Prof. Yeşim Aydın Son
Supervisor, **Health Informatics Dept., METU** _____

Examining Committee Members:

Assoc. Prof. Tunca Doğan
Computer Science & AI Engineering Dept., _____
Hacettepe University

Assoc. Prof. Yeşim Aydın Son
Health Informatics Dept., METU _____

Asst. Prof. Burçak Otlu
Health Informatics Dept., METU _____

Date: 25.01.2023

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name : Hüseyin Cahit Burduroğlu

Signature : _____

ABSTRACT

COMPARATIVE ANALYSIS OF BRAIN CELL CULTURES AND TISSUES IN ALZHEIMER'S DISEASE BASED ON DIFFERENTIAL EXPRESSION AND GENE SET ENRICHMENT

Burduroğlu, Hüseyin Cahit

MSc., Department of Health Informatics

Supervisor: Assoc. Prof. Yeşim Aydın Son

January 2023, 200 pages

Alzheimer's disease is currently the most common cause of dementia in the world. It is a neurodegenerative disease that is diagnosed neuropathologically by observing B-amyloid plaques and neurofibrillary tangles in the brain. Transcriptional differentiations, protein regulations, and the interactions in between have been investigated by recent studies to understand from which brain cell type the disease stems, such as microglia, astrocytes, and neurons. These studies are mostly performed on brain tissue samples that bring certain challenges. With the addition of novel gene editing methods in recent years, cell models are gaining popularity as an alternative to studying Alzheimer's disease. Yet, due to the complicated nature of the disease, cell models may not be representative enough. This study aims to determine the representativeness of cell models, in contrast to brain tissue samples, by comparing the two groups by their differentially expressed genes and enriched cellular functions. For this purpose, 19 projects (10 tissues, 9 cells) were selected for analysis from NCBI BioProject, which were then grouped by sex and cell types. The comparisons were made using the number of genes in gene sets that enrich certain biological processes. Distance matrices and topology plots were used for visualization.

Keywords: Alzheimer's disease, transcriptomics, functional enrichment, cell models

ÖZ

ALZHEIMER HASTALIĞINDA BEYİN HÜCRESİ KÜLTÜRLERİ VE DOKULARIN FARKLI GEN ANLATIMI VE GEN SETİ ZENGİNLEŞTİRME YÖNTEMLERİYLE KARŞILAŞTIRMALI ANALİZİ

Burduroğlu, Hüseyin Cahit

Yüksek Lisans, Sağlık Bilişimi Bölümü

Tez Yöneticisi: Doç. Dr. Yeşim Aydın Son

Ocak 2023, 200 sayfa

Alzheimer hastalığı dünyada bunamanın en büyük sebebi olarak bilinmektedir. Amiloyid beta plaklarının ve Tau proteinlerinden oluşan ipliklerin beyin dokusunda birikmesi sonucu gözlenen nörodejeneratif bir hastalıktır. Transkripsiyonel farklılıklar, protein anlatımı düzenlenmeleri ve bunların etkileşimleri yakın dönemdeki çalışmalarda araştırılarak hastalığın köken aldığı hücre tipleri anlaşılmaya çalışılmaktadır. Bu araştırmalar doku örnekleri üzerinde yapılmaktadır, ki bu örnekler beraberinde pek çok zorluk getirmektedir. Bunların üstesinden gelmek için, yakın dönemde geliştirilen gen düzenleme metotlarının da yardımıyla, hücre modellerinin kullanımı yaygınlaşmaktadır. Fakat, çevresel faktörlerden etkilenen ve sebebi tam olarak anlaşılamayan Alzheimer hastalığında hücre modellerinin doku örneklerini yansıtma düzeyi bir tartışma konusudur. Bu çalışma, hücre modellerinin doku örneklerini ne kadar temsil ettiğini tanımlama amacı taşımaktadır. Bu amaçla, NCBI BioProject üzerinde 19 adet çalışma elde edilmiş ve bu çalışmalar cinsiyet ve hücre tiplerine göre gruplandırılarak, DEG ve GSE analizleri üzerinden karşılaştırılmıştır. Karşılaştırmalar, herhangi bir biyolojik işleme denk düşen gen setlerindeki gen sayıları üzerinden yapılmış, uzaklık matriksleri ve topoloji grafikleri üzerinden görselleştirilmiştir.

Anahtar Sözcükler: Alzheimer hastalığı, transkriptomik, fonksiyonel zenginleştirme, hücre modelleri

To My Family

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my supervisor, Assoc. Prof. Yeşim Aydın Son, for her guidance and moral support. I was part of multiple studies while I was working on my thesis, and I've always felt her trust in me, which was encouraging.

Besides my supervisor, I want to thank Asst. Prof. Burçak OTLU and Assoc. Prof. Tunca DOĞAN for sparing their time for my thesis defense, and their constructive feedbacks.

I would also like to thank my beloved wife. She helped me to take my first steps and find my path back when I was lost. If I was able to focus on this study, it was thanks to her.

TABLE OF CONTENTS

ABSTRACT iv
ÖZ v
DEDICATION vi
ACKNOWLEDGMENTS vii
TABLE OF CONTENTS viii
LIST OF TABLES xii
LIST OF FIGURES xiii
LIST OF ABBREVIATIONS xviii
CHAPTERS
INTRODUCTION 1
1.1. Motivation 1
1.2. Alzheimer's Disease 2
1.2.1. Progression and Pathology 3
1.2.2. Genetic Background 7
1.2.3. Brain Cell Types and Their Characteristics in AD 11
1.2.4. Prevention and Treatment 17
1.3. Sequencing Technologies 18
1.3.1. First-Generation Sequencing 20
1.3.2. Next-Generation Sequencing 21
1.3.3. Third-Generation Sequencing 28
MATERIALS AND METHODS 31
2.1. Gathering and Manual Filtering of the Raw Data 31
2.2. Data Processing 32
2.2.1. Acquiring the Data 33
2.2.2. Data Analysis and Cleaning 33
2.2.3. Acquiring the Reference Genome and Genome Annotations 33
2.2.4. Aligning the Reads to the Reference Genome 33
2.2.5. Strandness of the Experiments 34

2.2.6. Generating Count Matrices for Genes 34
2.3. Differentially Expressed Gene (DEG) Analysis 35
2.4. Gene Set Enrichment by Gene Ontology Terms 37
2.4.1. Bringing GSE Results Together 38
2.4.2. GO Annotations and Analysis 39
RESULTS 41
3.1. Differentially Expressed Genes Analysis 41
3.2. Gene Set Enrichment Analysis 44
3.3. Distance Matrix Analysis 46
3.4. GO Topology Analysis 47
3.4.1. Astrocytes 49
3.4.2. Microglia 50
3.4.3. Neurons 51
3.4.4. Comparisons of Tissue Groups 53
3.4.5. Comparison of Cell Types 54
DISCUSSION 59
4.1. Conclusion 62
4.2. Future Studies 62
REFERENCES 65
APPENDICES 83
APPENDIX A 83
7.1. Experimental Designs 83
7.1.1. PRJNA232669 83
7.1.2. PRJNA279526 84
7.1.3. PRJNA399530 84
7.1.4. PRJNA451437 85
7.1.5. PRJNA482601 86
7.1.6. PRJNA527202 86
7.1.7. PRJNA559812 87
7.1.8. PRJNA576835 87
7.1.9. PRJNA603192 88
7.1.10. PRJNA643561 89

7.1.11. PRJNA644383 90
7.1.12. PRJNA662330 91
7.1.13. PRJNA675864 91
7.1.14. PRJNA683625 92
7.1.15. PRJNA688060 92
7.1.16. PRJNA688885 92
7.1.17. PRJNA714081 93
7.1.18. PRJNA727602 93
7.1.19. PRJNA767074 95
7.2. Alignment and Annotation Results 95
APPENDIX B 97
8.1. PRJNA232669 97
8.2. PRJNA279526 98
8.3. PRJNA399530 103
8.4. PRJNA451437 106
8.5. PRJNA482601 107
8.6. PRJNA527202 109
8.7. PRJNA559812 113
8.8. PRJNA576835 114
8.9. PRJNA603192 117
8.10. PRJNA643561 123
8.11. PRJNA644383 127
8.12. PRJNA662330 129
8.13. PRJNA675864 131
8.14. PRJNA683625 133
8.15. PRJNA688060 135
8.16. PRJNA688885 137
8.17. PRJNA714081 139
8.18. PRJNA727602 140
8.19. PRJNA767074 144
APPENDIX C 147
9.1. Lists of GO Terms and Tree Plots 147

9.1.1. Intersecting GO Terms for All Tissues without Sex Variable, not Found in Cell Samples 147

9.1.2. Intersecting GO Terms for All Female Tissues, not Found in Cell Samples 154

9.1.3. Intersecting GO Terms for All Male Tissues, not Found in Cell Samples 158

9.1.4. Intersecting GO Terms Specific to Female Tissues 163

9.1.5. Intersecting GO Terms Specific to Male Tissues 165

9.1.6. Intersecting GO Terms for All Astrocyte Samples, not Found in Tissue Samples 167

9.1.7. Intersecting GO Terms for All Microglia Samples, not Found in Tissue Samples 169

9.1.8. Intersecting GO Terms for All Neuron Samples, not Found in Tissue Samples 171

9.1.9. Intersecting GO Terms Specific to Astrocyte Samples 173

9.1.10. Intersecting GO Terms Specific to Microglia Samples 174

9.1.11. Intersecting GO Terms Specific to Neuron Samples 176

9.1.12. Intersecting GO Terms found in Neuron and Microglia Samples, but not in Astrocyte Samples 178

9.2. Complete Distance Matrices 179

9.3. Topologies of All Samples and Comparisons 181

LIST OF TABLES

Table 1 – Braak stages for NFT distribution in CNS. 6
Table 2 – Braak stages for Aβ plaque distribution in CNS. 6
Table 3 – Thal stages for Aβ plaque distribution in CNS. 6
Table 4 – A partial example count matrix. 35
Table 5 – An example design matrix. 36
Table 6 – Categorization of GSE results 38
Table 7 – Partial BP counts table. 40
Table 8 – Number of significant genes by types, subtypes, and comparisons. 41
Table 9 – Design matrix of PRJNA232669 81
Table 10 – Design matrix of PRJNA279526 82
Table 11 - Design matrix of PRJNA399530 82
Table 12 - Design matrix of PRJNA451437 83
Table 13 - Design matrix of PRJNA482601 84
Table 14 - Design matrix of PRJNA527202 84
Table 15 - Design matrix of PRJNA559812 85
Table 16 - Design matrix of PRJNA576835 85
Table 17 - Design matrix of PRJNA603192 86
Table 18 - Design matrix of PRJNA643561 87
Table 19 - Design matrix of PRJNA644383 88
Table 20 - Design matrix of PRJNA662330 89
Table 21 - Design matrix of PRJNA675864 89
Table 22 - Design matrix of PRJNA683625 90
Table 23 - Design matrix of PRJNA688060 90
Table 24 - Design matrix of PRJNA688885 90
Table 25 - Design matrix of PRJNA714081 91
Table 26 - Design matrix of PRJNA727602 92
Table 27 - Design matrix of PRJNA767074 93
Table 28 - Alignment and assignment results of trimmed data 93
Table 29 - Enriched GO terms found in tissues without sex variable and not in cells 145
Table 30 - Enriched GO terms found in female tissues and not in cells 152
Table 31 - Enriched GO terms found in male tissues and not in cells 156
Table 32 - Enriched GO terms only found in female tissues. 161
Table 33 - Enriched GO terms only found in male tissues. 163
Table 34 - Enriched GO terms found in astrocytes and not tissues. 165
Table 35 - Enriched GO terms found in microglia and not tissues 167
Table 36 - Enriched GO terms found in neuron and not tissues. 169
Table 37 - Enriched GO terms found only in astrocytes. 170
Table 38 - Enriched GO terms found only in microglia. 172
Table 39 - Enriched GO terms found only in neurons 174
Table 40 - Enriched GO terms tree plot common only in neurons and microglia 175

LIST OF FIGURES

Figure 1 – Number of PubMed search results for “AD cell model” over the years. 2
Figure 2 – Healthy and AD brain anatomy. 4
Figure 3 – Neuropathological indications observed on the macroscopic and microscopic levels in the AD brain. 5
Figure 4 - The distribution of NFTs over 6 Braak stages. 7
Figure 5 - Coronal, axial , and sagittal views of the brain showing the distribution of Aβ Plaques. 7
Figure 6 – Amyloid cascade hypothesis. 8
Figure 7 - 3D cartoon visualization of amyloid fibrillations. 9
Figure 8 – Reactive astrocytes. 13
Figure 9 - Astrocyte and microglia in their normal and reactive stages 15
Figure 10 – Oligodendrocyte forming myelin sheath over axons 16
Figure 11 – Risk factors for AD 18
Figure 12 – PCR steps. 20
Figure 13 – Resulting bands from a Sanger sequencing experiment. 21
Figure 14 – Current NGS techniques. 22
Figure 15 – Example of a FastQ format. 23
Figure 16 – An example analysis of a FastQ file with FastQC. 24
Figure 17 - An example adapter removal process command with cutadapt. 24
Figure 18 – An example of two BAM files shown in IGV. 26
Figure 19 – An example graph from the “Sequence Duplication Levels” module of FastQC. 27
Figure 20 – PacBio SMRT sequencing. 29
Figure 21 – Oxford Nanopore sequencing. 30
Figure 22 – The list of projects that were used in the downstream analysis. 32
Figure 23 – The workflow of BAM generation. 34
Figure 24 – The steps for acquiring the strandness information and count matrix generation. 35
Figure 25 – An example DESeq2 workflow. 36
Figure 26 – Usage of gseGO() function from the clusterProfiler package. 39
Figure 27 – BP topology graph of astrocyte, given as an example. 40
Figure 28 – Percentage of significant genes by sample group. 42
Figure 29 – Volcano plots of PRJNA603192 samples given as an example. 43
Figure 30 – Number of enriched GO terms by samples in each project. 45
Figure 31 – Distance matrix of cells vs. brain regions by BP. 46
Figure 32 – Distance matrix of cell samples vs. tissues by the number of genes in each BP term. 47
Figure 33 – BP topologies of major sample groups. 48
Figure 34 – MF topologies of major sample groups. 48
Figure 35 – CC topologies of major sample groups. 49

Figure 36 – Astrocyte topologies in comparison to tissues. 50
Figure 37 – Microglia topologies in comparison with tissues. 51
Figure 38 – Neuron topologies in comparison with tissues. 51
Figure 39 - Distance matrix of cell models by BP. 52
Figure 40 - Distance matrix of brain regions by BP. 53
Figure 41 – Distance matrix of brain regions vs. sexes by BP. 54
Figure 42 - Biological processes that are affected in AD, specific to astrocytes. 55
Figure 43 - Biological processes that are affected in AD, specific to microglia. 56
Figure 44 - Biological processes that are affected in AD, specific to neurons. 57
Figure 45 – The common biological processes that are affected in AD, in both microglia and neurons. 58
Figure 46 – Biological processes that are affected in AD, specific to the female samples. 60
Figure 47 – Biological processes affected in AD, specific to male samples. 61
Figure 48 – Summary graphs of PRJNA232669. 95
Figure 49 - Summary graphs of PRJNA279526 without sex variable. 96
Figure 50 – Heatmap of PRJNA279526 samples. 97
Figure 51 – P-value distributon and volcano plot of female and male samples. 98
Figure 52 – PCA plot of PRJNA270529, considering sex variable. 99
Figure 53 – Heatmap of PRJNA279526. 99
Figure 54 – P-value distribution by brain cell types. 100
Figure 55 – Volcano plots of brain cell types. 101
Figure 56 – Distance matrix of PRJNA399530. 102
Figure 57 - PCA plot of PRJNA399530. 102
Figure 58 - Heatmap of PRJNA399530. 103
Figure 59 – Summary graphs of PRJNA451437. 104
Figure 60 – Summary graphs of PRJNA482601. 105
Figure 61 - Heatmap of PRJNA482601. 106
Figure 62 – P-value distributions of PRJNA527202. 107
Figure 63 - Volcano plots of PRJNA527202. 108
Figure 64 - Distance matrix of PRJNA527202. 109
Figure 65 - PCA plot of PRJNA527202. 109
Figure 66 - Heatmap of PRJNA527202. 110
Figure 67 - Summary graphs of PRJNA559812. 111
Figure 68 - Summary graphs of PRJNA576835. 112
Figure 69 - P-value distributions and volcano plots by sex. 113
Figure 70 - PCA plot of PRJNA576835. 114
Figure 71 - Heatmap of PRJNA576835. 114
Figure 72 - Distance matrix of PRJNA603192. 115
Figure 73 - PCA plot of PRJNA603192. 116
Figure 74 - Heatmap of PRJAN603192. 117
Figure 75 - P-value distributions and volcano plots of samples without considering sex variable. 118
Figure 76 - P-value distributions and volcano plots of female samples. 119
Figure 77 - P-value distributions and volcano plots of male samples. 120

Figure 78 - Distance matrix of PRJNA643561. 121
Figure 79 – PCA plot of PRJNA643561. 122
Figure 80 - Heatmap of PRJNA643561. 122
Figure 81 - P-value distribution and volcano plots of SORL1 mutants. 123
Figure 82 - P-value distributions and volcano plots of TREM2 mutants. 124
Figure 83 - Summary graphs of PRJNA644383. 125
Figure 84 - Heatmap of PRJNA644383. 126
Figure 85 - Summary graphs of PRJNA662330. 127
Figure 86 - Heatmap of PRJNA662330. 128
Figure 87 - summary graphs of PRJNA675864. 129
Figure 88 – Heatmap of PRJNA675864. 130
Figure 89 - Summary graphs of PRJNA682625. 131
Figure 90 - Heatmap of PRJNA683625. 132
Figure 91 - Summary graphs of PRJNA688060. 133
Figure 92 - Heathmap of PRJNA688060. 134
Figure 93 - Summary graphs of PRJNA688885. 135
Figure 94 - Heatmap of PRJNA688885. 136
Figure 95 - Summary graphs of PRJNA714081. 137
Figure 96 - Summary graphs for PRJNA727602 without considering sex variable. 138
Figure 97 - P-value distribution and volcano plots of female and male samples. 139
Figure 98 - PCA plot of PRJNA727602. 140
Figure 99 - Heatmap of PRJNA727602. 141
Figure 100 - Summary graphs of PRJNA767074. 142
Figure 101 - Heatmap of PRJNA767074. 143
Figure 102 – Enriched BP terms tree plot found in tissues without sex variable and not in cells. 144
Figure 103 – Enriched CC terms tree plot found in tissues without sex variable and not in cells 145
Figure 104 – Enriched MF terms tree plot found in tissues without sex variable and not in cells 145
Figure 105 – Enriched BP terms tree plot foound in female tissues and not in cells 151
Figure 106 – Enriched CC terms tree plot found in female tissues and not in cells 151
Figure 107 – Enriched MF terms tree plot found in female tissues and not in cells 152
Figure 108 – Enriched BP terms tree plot found in male tissues and not in cells 155
Figure 109 – Enriched CC terms tree plot found male tissues and not in cells 156
Figure 110 – Enriched MF terms tree plot found in male tissues and not in cells 156
Figure 111 – Enriched BP terms tree plot only found in female tissues. 160
Figure 112 – Enriched CC terms tree plot only found in female tissues. 160
Figure 113 – Enriched MF terms tree plot only found in female tissues. 161
Figure 114 – Enriched BP terms tree plot only found in male tissues. 162
Figure 115 – Enriched CC terms tree plot only found in male tissues. 162
Figure 116 – Enriched MF terms tree plot only found in male tissues. 163
Figure 117 – Enriched BP terms tree plot found in astrocytes and not tissues. 164
Figure 118 – Enriched CC terms tree plot found in astrocytes and not tissues. 165

Figure 119 – Enriched BP terms tree plot found in microglia and not tissues. 166
Figure 120 – Enriched CC terms tree plot found in microglia and not tissues. 166
Figure 121 – Enriched MF terms tree plot found in microglia and not tissues. 167
Figure 122 – Enriched BP terms tree plot found in neuron and not tissues..... 168
Figure 123 – Enriched CC terms tree plot found in neuron and not tissues..... 169
Figure 124 – Enriched MF terms tree plot found in neuron and not tissues..... 169
Figure 125 – Enriched BP terms tree plot found only in astrocytes. 170
Figure 126 – Enriched BP terms tree plot found only in microglia. 171
Figure 127 – Enriched CC terms tree plot found only in microglia. 171
Figure 128 – Enriched MF terms tree plot found only in microglia. 172
Figure 129 – Enriched BP terms tree plot found only in neurons. 173
Figure 130 – Enriched CC terms tree plot found only in neurons 173
Figure 131 – Enriched MF terms tree plot found only in neurons 174
Figure 132 – Enriched BP terms tree plot common only in neurons and microglia 175
Figure 133 – Enriched MF terms tree plot common only in neurons and microglia 175
Figure 134 – Distance matrix of all sample groups, built upon BP 176
Figure 135 - Distance matrix of all sample groups, built upon CC 177
Figure 136 – Distance matrix of all sample groups, built upon MF 178
Figure 137 – BP topologies and comparisons of tissues grouped by sex 178
Figure 138 - MF topologies and comparisons of tissues grouped by sex 179
Figure 139 - CC topologies and comparisons of tissues grouped by sex 179
Figure 140 - BP topologies and comparisons of astrocytes vs. tissues grouped by sex 180
Figure 141 - MF topologies and comparisons of astrocytes vs. tissues grouped by sex 180
Figure 142 - CC topologies and comparisons of astrocytes vs. tissues grouped by sex 181
Figure 143 - BP topologies and comparisons of microglia vs. tissues grouped by sex 181
Figure 144 - MF topologies and comparisons of microglia vs. tissues grouped by sex 182
Figure 145 - CC topologies and comparisons of microglia vs. tissues grouped by sex 182
Figure 146 - BP topologies and comparisons of neurons vs. tissues grouped by sex 183
Figure 147 - MF topologies and comparisons of neurons vs. tissues grouped by sex 183
Figure 148 - CC topologies and comparisons of neurons vs. tissues grouped by sex 184
Figure 149 - BP topologies and comparisons of neural progenitors vs. tissues grouped by sex 184
Figure 150 - MF topologies and comparisons of neural progenitors vs. tissues grouped by sex 185

Figure 151 - CC topologies and comparisons of neural progenitors vs. tissues grouped by sex 185
Figure 152 - BP topologies and comparisons of cerebral organoids vs. tissues grouped by sex 186
Figure 153 - MF topologies and comparisons of cerebral organoids vs. tissues grouped by sex 186
Figure 154 - CC topologies and comparisons of cerebral organoids vs. tissues grouped by sex 187
Figure 155 - BP topologies and comparisons of iPSCs vs. tissues grouped by sex .. 187
Figure 156 - MF topologies and comparisons of iPSCs vs. tissues grouped by sex . 188
Figure 157 - CC topologies and comparisons of iPSCs vs. tissues grouped by sex .. 188
Figure 158 - BP topologies and comparisons of astrocytes vs. other cells 189
Figure 159 - MF topologies and comparisons of astrocytes vs. other cells 189
Figure 160 - CC topologies and comparisons of astrocytes vs. other cells 190
Figure 161 - BP topologies and comparisons of microglia vs. other cells 190
Figure 162 - MF topologies and comparisons of microglia vs. other cells 191
Figure 163 - CC topologies and comparisons of microglia vs. other cells 191
Figure 164 - BP topologies and comparisons of neurons vs. other cells 192
Figure 165 - MF topologies and comparisons of neurons vs. other cells 192
Figure 166 - CC topologies and comparisons of neurons vs. other cells 193
Figure 167 - BP topologies and comparisons of neural progenitors vs. other cells 193
Figure 168 - MF topologies and comparisons of neural progenitors vs. other cells 194
Figure 169 - CC topologies and comparisons of neural progenitors vs. other cells 194
Figure 170 - BP topologies and comparisons of cerebral organoids vs. other cells 195
Figure 171 - MF topologies and comparisons of cerebral organoids vs. other cells 195
Figure 172 - CC topologies and comparisons of cerebral organoids vs. other cells 196
Figure 173 - BP topologies and comparisons of iPSCs vs. other cells 196
Figure 174 - MF topologies and comparisons of iPSCs vs. other cells 197
Figure 175 - CC topologies and comparisons of iPSCs vs. other cells 197

LIST OF ABBREVIATIONS

AD	Alzheimer's disease
ADAD	Autosomal dominant Alzheimer's disease
ADAS	Alzheimer's disease assessment scale
ADAS-Cog	ADAS cognitive subscale
APOE	Apolipoprotein E
APP	Amyloid precursor protein
ast	Astrocyte
ATP	Adenozine triphosphate
Aβ	Amyloid beta
bai	BAM index
BAM	Binary alignment map
BED	Browser extensible data
BLD	Blood
BP	Biological process
CAA	Cerebral amyloid angiopathy
CC	Cellular component
CDR-SOB	Clinical dementia rating scale
cer	Cerebral organoid
CNS	Central nervous system
CPU	Central processing unit
CSF	Cerebrospinal fluid
ddATP	di-deoxyadenozine triphosphate
ddNTP	di-deoxynucleotide triphosphate
DEG	Differentially expressed genes
DNA	Deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
DPC	Dorsolateral prefrontal cortex
dsDNA	Double stranded DNA
EOAD	Early-onset Alzheimer's disease
fem	Female
GB	Gigabyte
GEO	Gene Expression Omnibus
GO	Gene ontology
GSE	Gene set enrichment
GTF	Gene transfer format
GWAS	Genome-wide association studies
hESC	Human embryonic stem cell
HPC	Hippocampus
IGV	Integrative Genomics Viewer
IPL	Inferior parietal lobule

iPSC	Induced pluripotent stem cell
LOAD	Late-onset Alzheimer's disease
LTL	Lateral temporal lobe
mal	Male
MF	Molecular function
mgl	Microglia
Mhz	Megahertz
MMSE	Mini-mental state examination
mRNA	Messenger RNA
MTG	Middle temporal gyrus
NCBI	National Center for Biotechnology Information
neu	Neuron
NFT	Neurofibrillary tangle
NGS	Next-generation sequencing
npg	Neural progenitor
OLF	Olfactory bulb
ONT	Oxford Nanopore Technologies
OPC	Oligodendrocyte precursor cells
PacBio	Pacific Biosciences
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PSEN1	Presenilin 1
PSEN2	Presenilin 2
RAM	Random access memory
RNA	Ribonucleic acid
RNAseq	RNA sequencing
SAM	Sequence alignment map
scRNAseq	Single cell RNA sequencing
SMRT	Single molecule real-time
SNP	Single nucleotide polymorphism
ssDNA	Single stranded DNA
stc	Induced pluripotent stem cell
Tm	Melting temperature
UTR	Untranslated region
VCF	Variant call format
WES	Whole exome sequencing
WGS	Whole genome sequencing
ZMW	Zero-mode waveguide

CHAPTER 1

INTRODUCTION

1.1. Motivation

Alzheimer's disease (AD) is a condition that progresses over an individual's lifetime without showing symptoms until the age of 60. Even though there are known underlying genetic conditions, these cannot be diagnosed by only sequencing the individual's DNA and looking for high-risk variants related to AD. The disease is characterized by senile plaque and neurofibrillary tangle formations in the brain tissue (Goedert & Spillantini, 2006). Neuropathological, transcriptomic, and proteomic analyses can only be done after the individual passes away. These analyses can be performed on post-mortem tissue which brings challenges and limitations.

The first challenge is in the area of ethics. There are administrative rules for the usage of human tissues for research, and the researcher must have approval for their study.

After the individual passes away, there are several steps to gather, contain and preserve the tissues. In addition to factors released in all tissues of the organism, these steps affect the transcriptomic profile of the tissue of interest, reducing the overall quality and, ultimately, the results. Preservation methods do not guarantee the protection of genomic molecules either; these molecules can be damaged or degraded over time (Gomez-Nicola & Boche, 2015).

The studies searching for candidate genes also require healthy samples as a reference to compare with the findings from the AD samples. When this is the case, AD and healthy samples must be age and sex-matched. The post-mortem interval, cause of death, and medications used by the passed-away individual also have to be considered for the sample to be used in the study (Gomez-Nicola & Boche, 2015).

Because of all these restrictions on working on human tissues, animal and cell models can be good alternatives depending on the disease worked on. Cell models also provide the advantage of being less costly and letting the researcher have a direct influence on the cells, such as adding compounds without considering the metabolic processes of the total organism or being able to work on a single type of cell of a tissue with multiple cell types (Cetin et al., 2022). With recent developments in the area of gene editing, such as CRISPR Cas-9, it is possible to

work with AD-related variants on cell models (Redman et al., 2016). Induced pluripotent stem cells (iPSC) or human embryonic stem cells (hESC) (T. Liu et al., 2020a) are gene-edited to carry variants that increase the risk of AD. Then, brain cells (neurons and glia) are generated from these stem cells. This is one way to obtain AD-like brain cell cultures (Dolmetsch & Geschwind, 2011). Another way is to expose these cells to the extracellular environment observed in AD brains (Chen et al., 2021a). However, AD is a condition with many underlying reasons, most of which are not clearly understood. These models are generated by the statistical information gathered from the genome-wide association studies that can only show candidate genes and single nucleotide polymorphisms (SNPs). It is a known issue that AD is caused not only by these variants but also by many environmental factors (Gatz et al., 2006; Scheltens et al., 2016).

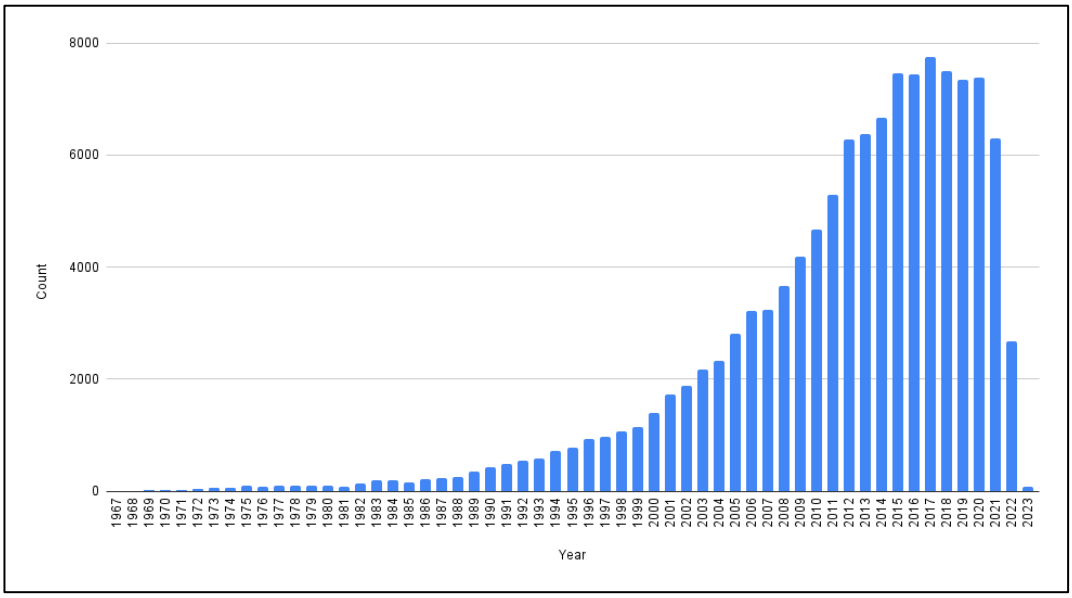

Figure 1 – Number of PubMed search results for "AD cell model" over the years.

This study aims to understand the similarities and differences between AD-like cell cultures and AD tissue samples in terms of transcriptomic profiles and to answer if the cell cultures can be used for AD modeling, mainly from a transcriptomic perspective.

1.2. Alzheimer's Disease

First defined on November 3rd, 1906, Alzheimer's disease (AD) is the most common neurodegenerative disease and the leading cause of cognitive impairment or dementia (Goedert & Spillantini, 2006; Scheltens et al., 2016). In the US, only 4.5 million people were diagnosed as AD patients in 2000, and the number is expected to be quadrupled by 2050 (Hebert et al., 2003). According to another study, the global number is predicted as 24 million in 2012 and is expected to be doubled every 20 years (Mayeux & Stern, 2012). It is also the sixth leading reason of death in the US (Atri, 2019).

The majority of the cases are diagnosed after the age of 65, which are identified as "late-onset AD (LOAD)," while less than 5% of the cases occur before 65, termed

"early-onset AD (EOAD)." Another type of the disease can occur with 1% odds and be diagnosed at very early ages, which can progress rapidly. This type comes from an autosomal dominant inheritance, termed "autosomal dominant Alzheimer's disease" (ADAD) (Long & Holtzman, 2019).

1.2.1. Progression and Pathology

AD is a disease that slowly progresses over the years. The symptoms are observed when the disease enters the late stages, and behavioral changes occur. Observations of behavior and cognitive testing are used during diagnosis. Macroscopic and microscopic pathology of the tissue can only be observed post-mortem after the individual passes away.

1.2.1.1 Behavioral Pathology

The earlier symptoms of the disease are characterized by learning and memory impairments. As the disease progresses, these symptoms are followed by further impairments in attention, language, visuospatial functions, and behavioral abnormalities (McKhann et al., 2011). The progression speed varies between the type of the disease, the individual's genetic background, and the environmental factors (Wainaina et al., 2014).

During the diagnosis process, a few critical features of AD-related dementia is considered to separate it from dementia from another cause. Dementia in AD patients progresses over time instead of occurring suddenly, as in hours or days. The condition worsens with progression; it does not stay the same over the years or heal. Although not very common, amnesia may not be the identifying symptom for some patients. Hence, in terms of diagnosing AD through dementia, the examination is separated into two categories: Amnestic presentation and non-amnestic presentation. Patients showing amnestic presentation have trouble with remembering recent information, while patients showing non-amnestic presentation have trouble with finding the correct words during speech (lingual dysfunction), recognizing faces, objects, written words, color discrimination (visuospatial presentation), impaired problem-solving and judgment (executive dysfunction) (Boedeker et al., 2022; McKhann et al., 2011).

Other common behavioral symptoms of AD include irritability, agitation, delusions, depression, and anxiety, which are episodic and non-cognitive symptoms. The occurrence of these symptoms increases as the disease progresses. With progression, hallucinations, delirium, and personality changes may follow. These symptoms also have varying degrees and can occur simultaneously or separately (Grossberg, 2002; Weiner et al., 2005). Some methods are developed to determine the progression level of the disease by considering the severity of these observable and behavioral symptoms, such as Alzheimer's Disease Assessment Scale (ADAS) – Cognitive Subscale (ADAS-Cog), Mini-Mental State Examination (MMSE), Clinical Dementia Rating scale (CDR-SOB). While ADAS-Cog and CDR-SOB are considered more accurate for identifying cognitive impairment levels than MMSE, ADAS-Cog remains the gold standard for measurements in patients with AD (Balsis et al., 2015; Mohs, 1996; Morris, 1997; Tombaugh & McIntyre, 1992).

1.2.1.2 Neuropathology

AD is highly associated with neurofibrillary tangles found within the cytoplasm of some neurons. The hyperphosphorylation of Tau proteins forms these tangles. Another mainly associated lesion with the disease is amyloid plaques (also called senile plaques) formed around beta-amyloids (Perl, 2010).

There are other lesions besides neurofibrillary tangles and senile plaques: Granulovacuolar degenerations, rod-like bodies termed Hirano bodies, and cerebral amyloid angiopathy. These are poorly understood lesions, as AD is related to many age-related processes (Perl, 2010).

In contrast to a healthy brain, the AD brain contains cortical atrophy in at least moderate amounts. The sulcal spaces are enlarged due to the atrophy, primarily visible on the frontal and temporal regions. The ventricles enlarge, the hippocampus shrinks, the volume of the white matter (Piguet et al., 2009), and the brain weight and neuronal pigmentation decrease. It is important to note that none of these alterations are specific to Alzheimer's disease (Deture & Dickson, 2019).

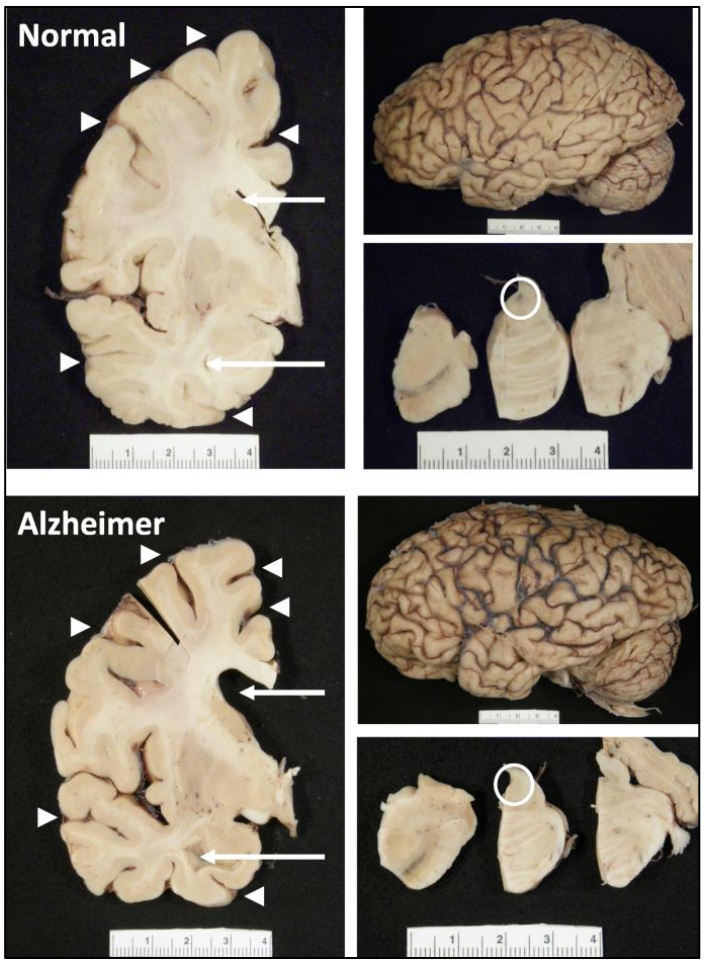

Figure 2 – Healthy and AD brain anatomy. Arrowheads show coronal sections, and arrows indicate ventricles. Due to atrophy, the spaces are enlarged. The circle shows neural pigmentation loss in the locus voeruleus (Deture & Dickson, 2019).

On the microscopic side, the beforementioned extracellular amyloid-beta (Aβ) plaques and intracellular neurofibrillary tangles (NFT) are observed. As these lesions accumulate over time in the brain, neurons phase into cell death, and synaptic structures start to degenerate, ultimately leading to behavioral and macroscopic symptoms. Aβ plaques disrupt the cell functions vital for the cells, and NFTs disrupt synaptic transport and communication between the cells (Deture & Dickson, 2019).

The Aβ plaques are separated into two categories by their core types: Dense-core and diffuse-core. Dense core plaques are the reason for synaptic loss leading to impairment of cognitive functions. They are surrounded by activated microglia and astrocytes. In contrast to the fibrillary structure of dense-core plaques, diffuse-core plaques are amorphous. They can naturally occur by age; hence they are not considered for AD diagnosis (Perl, 2010).

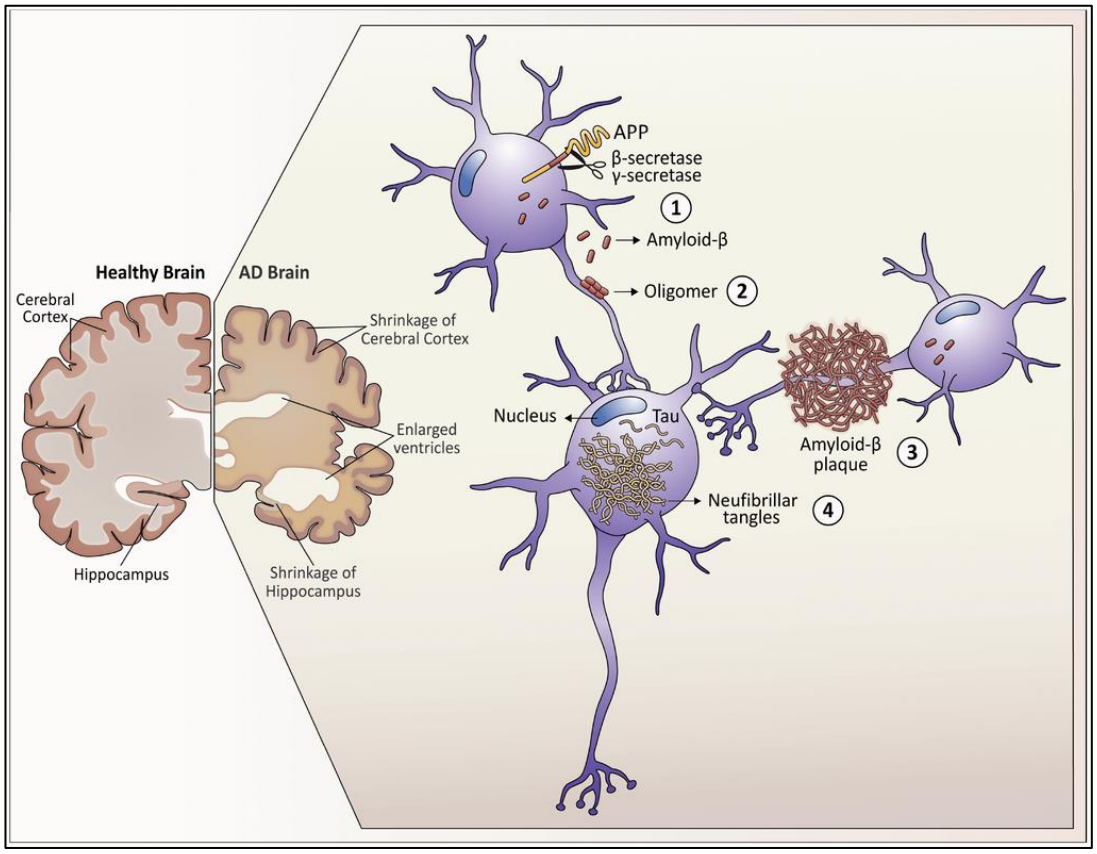

Figure 3 – Neuropathological indications observed on the macroscopic and microscopic levels in the AD brain. The ventricles are enlarged, and the hippocampus has shrunk. There are NFTs shown inside neurons, and amyloid plaques are formed in the extracellular matrix around the axons (Gomez et al., 2020).

Under normal conditions, Tau proteins are positioned within axons and responsible for microtubule stability. NFTs form in the neural cytoplasm when Tau proteins are hyperphosphorylated and misfolded. They can also be found in the extracellular matrix when a neuron dies and the membrane disintegrates (Perl, 2010).

The distribution of Aβ plaques and NFTs intensifies as the disease progresses. The stages of accumulation and distribution of NFTs are separated into six stages defined by Braak and Braak. For Aβ plaques, there are three stages defined by Braak and Braak, and five stages defined by Thal et al. These stages are termed Braak Stages and Thal Stages (Braak & Braak, 1991; Thal et al., 2002).

Table 1 – Braak stages for NFT distribution in CNS. (Braak & Braak, 1991)

Braak Stage	Distribution of NFTs
Stage I	trans entorhinal (perirhinal) region, entorhinal cortex
Stage II	hippocampus
Stage III	subiculum of the hippocampal formation
Stage IV	amygdala, thalamus, and claustrum
Stage V	Associative isocortical areas
Stage VI	primary sensory, motor, and visual areas

Table 2 – Braak stages for Aβ plaque distribution in CNS. (Braak & Braak, 1991)

Braak Stage	Distribution of Aβ Plaques
Stage A	basal portions of the frontal, temporal, and occipital lobes
Stage B	Associative isocortical areas, hippocampus
Stage C	primary isocortical areas, striatum, thalamus, hypothalamus, subthalamic nucleus, and red nucleus

Table 3 – Thal stages for Aβ plaque distribution in CNS. (Thal et al., 2002)

Thal Stage	Distribution of Aβ Plaques
Stage 1	Isocortical areas
Stage 2	the entorhinal cortex, hippocampal formation, amygdala, insular, and cingulated cortices
Stage 3	The striatum, basal forebrain cholinergic nuclei, thalamus and hypothalamus, and white matter
Stage 4	red nucleus, substantia nigra, reticular formation of the medulla oblongata, superior and inferior colliculi
Stage 5	The reticular formation, raphe nuclei, locus ceruleus, molecular layer of the cerebellum

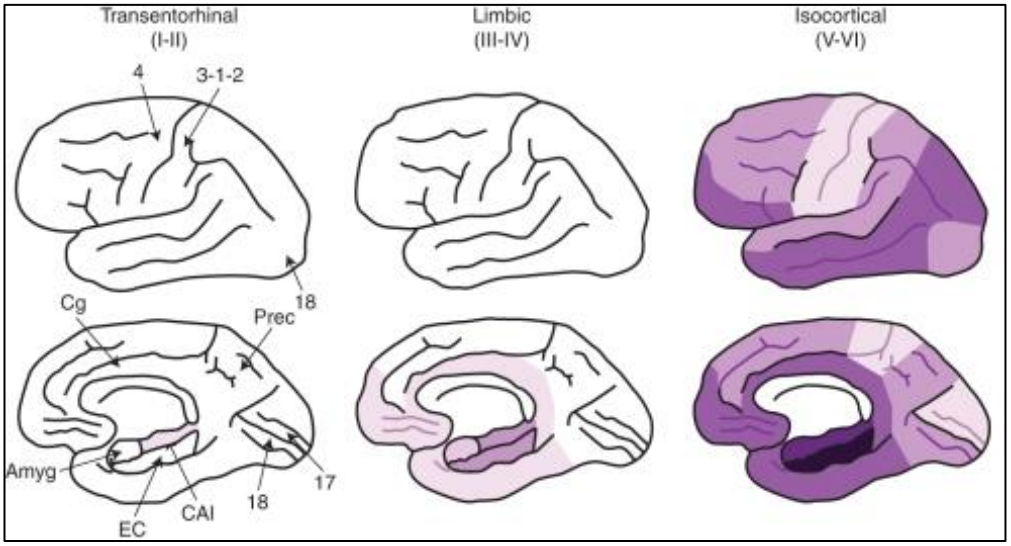

Figure 4 - The distribution of NFTs over 6 Braak stages. As the colors get darker, the intensity of NFTs increases. Amyg: Amygdala, EC: Entorhinal cortex, CA1: Cornus ammonis 1 hippocampal subfield, Cg: Cingulate cortex, Prec: Precuneus, 4: Primary motor cortex, 3-1-2: Primary sensory cortex, 17: Primary visual cortex, 18: Associative visual cortex (Braak & Braak, 1991; Serrano-Pozo et al., 2011)

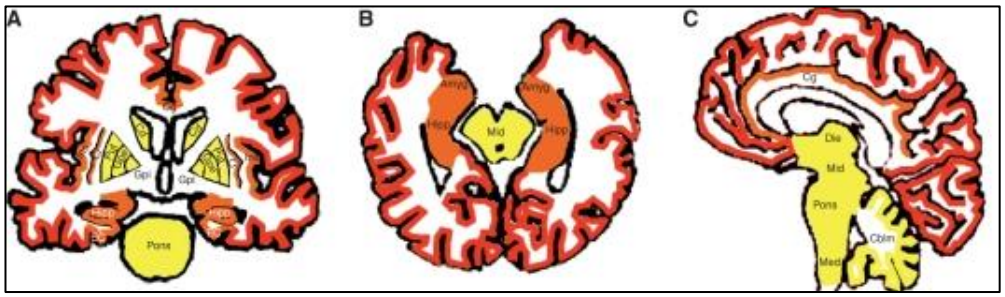

Figure 5 – Coronal (A), axial (B), and sagittal (C) views of the brain showing the distribution of Aβ Plaques. The progression starts with red areas, then orange, and lastly, yellow, according to Thal staging (Braak & Braak, 1991; Serrano-Pozo et al., 2011; Thal et al., 2002).

Aβ plaques can also accumulate in blood vessels in the brain and lead to cerebral amyloid angiopathy (CAA), which occurs in more than 85% of AD patients. CAA has two types: Type 1 – Arteries, arterioles, and capillaries are damaged, the patient usually has an ε4 variant of the *APOE* gene, Type 2 – Capillaries are not damaged, the patient usually has an ε2 variant of the *APOE* gene.

1.2.2. Genetic Background

The risk factor for AD significantly increases with familial history (Gatz et al., 2006).

Four genes play a critical role in the progression of AD. *APP*, *PSEN1*, and *PSEN2* are the genes where rare mutations lead to EOAD, and *APOE* is a gene with multiple variants where each variant carries a different amount of risk factors for the incidence of LOAD (Tanzi, 2012).

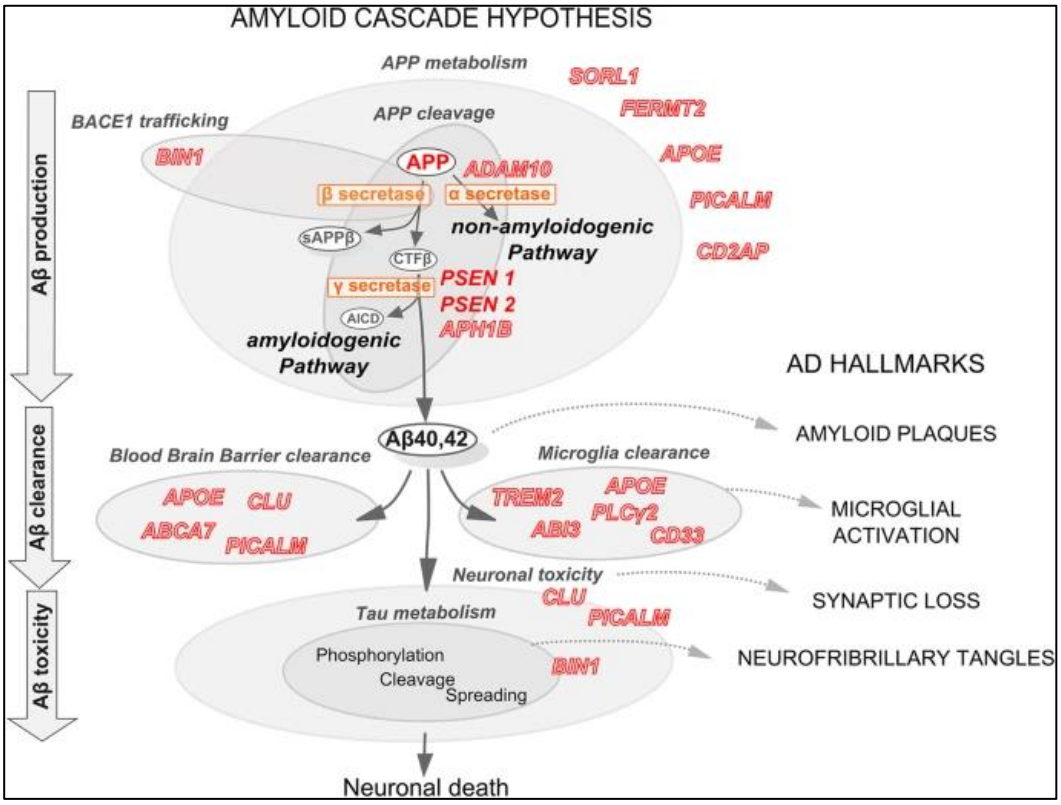

Figure 6 – Amyloid cascade hypothesis. Genes and interaction creating the amyloid cascade leading to AD. CLU, PICALM, and BIN1 can be seen in the Tau metabolism, while all other genes are related directly or indirectly to amyloid plaque formation (Dourlen et al., 2019).

1.2.2.1 APP

APP stands for "Amyloid β (A4) Protein Precursor". The gene is located on chromosome 21, locus 21. It expresses a protein found in neurons and glia positioned on the membrane (J. Wang et al., 2017). *APP* plays a role in synaptic repair, formation, and neural transportation (Priller et al., 2006; Satpute-Krishnan et al., 2006).

The synthesized precursor protein is cleaved by β-secretase and γ-secretase complex that can create Aβ40 and Aβ42 peptides, with a length of 40 and 42 amino acids, respectively. The two extra amino acids found in Aβ42 increases the amount of aggregation, according to studies on EOAD patients where there is an increased amount of Aβ42 (Goate et al., 1991).

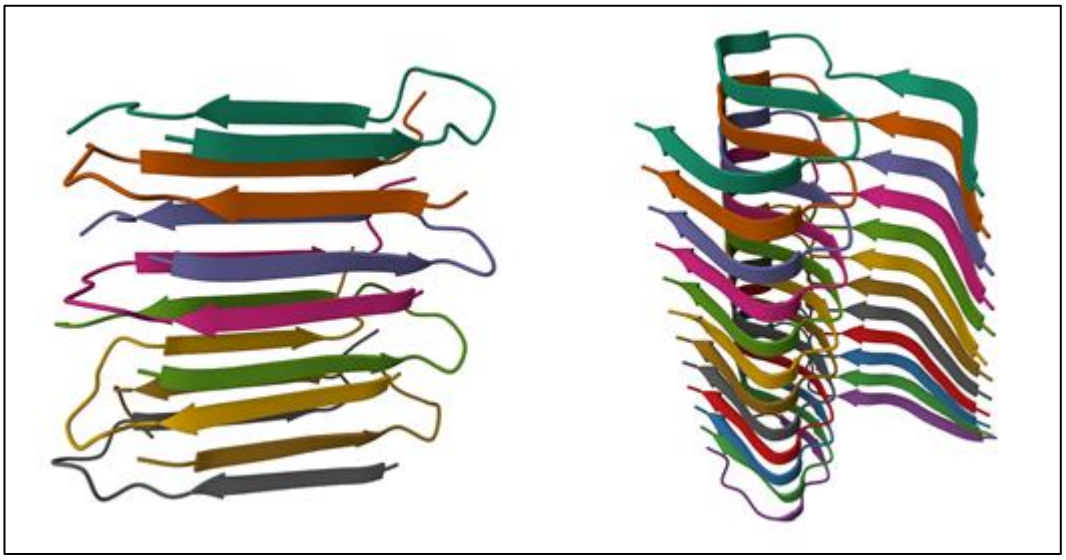

Figure 7 - 3D cartoon visualization of amyloid fibrillations. Fibrillary Aβ40 (on the left, PDB ID: 2LNQ) (Qiang et al., 2012) and Aβ42 (on the right, PDB ID: 2MXU) (Xiao et al., 2015) molecular structure. Every color represents a different chain of an amyloid peptide. Aβ40 carries an antiparallel architecture with flexible loops, while Aβ42 forms three beta strands directly in contact with the chain next to it. There are no flexible regions (H.M. Berman et al., 2021).

APP has 24 known mutations, including duplications that lead to increased Aβ42/ Aβ40 ratio, Aβ production, and aggregation (Tanzi, 2012).

1.2.2.2 PSEN1 & PSEN2

PSEN1 (presenilin 1) is positioned at chromosome 14, locus 24, while its homolog *PSEN2* (presenilin 2) is positioned at chromosome 1, locus 31. Amongst the mutation found within *APP*, *PSEN1*, and *PSEN2*, *PSEN1* is the most involved gene in the development of EOAD. *PSEN1* and *PSEN2* are responsible for producing presenilin that establishes the catalytic subunit of the γ-secretase complex, which cleaves the amyloid protein precursors to generate amyloid beta peptides (Lanoiselée et al., 2017).

There are 221 reported mutations for *PSEN1* and 19 reported for *PSEN2* related to AD (Lanoiselée et al., 2017). Mutations in these two genes increase Aβ42/Aβ40 ratio (Tanzi, 2012).

1.2.2.3 APOE

APOE stands for "Apolipoprotein E" and is responsible for transporting lipids, cholesterol, and lipid-soluble vitamins. It is positioned on chromosome 19, locus 13 (Tanzi, 2012). As transporting nutrients is the role of astrocytes, they express apoE most in the brain tissue. Microglia also express this protein, while neurons mainly synthesize apoE receptors (Z. Zhang et al., 2013). In addition to lipid metabolism, *APOE* also plays a role in the degradation of Aβ peptides; it enhances the degradation within microglia and extracellular matrix by making it easier for microglia to attach the plaques (Jiang et al., 2008).

There are three isoforms of the *APOE* protein, denoted as ε2, ε3, and ε4. ε3 is the most common form (Tanzi, 2012) and is considered the neutral type of *APOE* (C. C.

Liu et al., 2013). ε2 increases the protective effects of *APOE* against LOAD (Corder et al., 1994). Lastly, ε4 brings the highest amount of risk for developing LOAD. An ε4 carrier is four times more likely to develop the disease. The risk is more than 10-fold if the ε4 allele is homozygous (Tanzi, 2012).

APOE variants do not necessarily cause or protect the individual from AD. There are many environmental factors to be considered. Different demographics show different incidences. For example, ε4 frequency in Nigerian people is the highest in the world, but the incidence of AD is low in the population. It is believed that the underlying reason is the low cholesterol levels in their diet (Sepehrnia et al., 1989).

1.2.2.4 Other Genes

Identification of other gene candidates for the development of AD is made through genome-wide association studies (GWAS). In GWAS, known SNPs are tested for their relation to AD using samples from healthy individuals and individuals diagnosed with AD. Currently, there are 38 known loci associated with AD (Wightman et al., 2021). Some gene candidates that play a role in Aβ peptide accumulation (Tanzi, 2012) are listed below.

- *CD33*: Expresses a transmembrane protein that binds to sialic acids. The protein is found on the cells stem from the myeloid tissue. It plays a role in regulating the innate immune system (Garnache-Ottou et al., 2005).
- *CLU*: Also known as *APOJ*, *CLU* is responsible for apoptosis and clearance of cell debris (Jones & Jomary, 2002). It is believed to take a role in Aβ transport and formation of Aβ fibrils (DeMattos et al., 2004; Nuutinen et al., 2009).
- *CR1*: *CR1* expresses a membrane protein that regulates complement activation in some immune cells. It plays a role in reducing Aβ neurotoxicity (Lambert et al., 2009).
- *PICALM*: This gene plays a role in the cleavage of *APP* to Aβ by γ-secretase (Harold et al., 2009).
- *ABCA7*: A protein that is a member of the *ABC* family. This family is responsible for molecule transportation through the membrane (*ABCA7 ATP Binding Cassette Subfamily A Member 7 [Homo Sapiens (Human)] - Gene - NCBI*, n.d.). The gene plays a role in lipid metabolism and cellular signaling (Hollingworth et al., 2011).
- *CD2AP*: *CD2AP* expresses a structural molecule that plays a role in regulating the actin cytoskeleton. It also impacts cellular signaling (Hollingworth et al., 2011).
- *EPHA1*: This gene encodes a receptor that mediates nervous system development. Its roles contain cellular signaling and innate immunity (*EPHA1 EPH Receptor A1 [Homo Sapiens (Human)] - Gene - NCBI*, n.d.; Hollingworth et al., 2011).
- *MS4A6A*/*MS4A4E*: This gene encodes a transmembrane protein that plays a role in cellular signaling (Hollingworth et al., 2011).

- *ATXN1*: *ATXN1* is a gene that is conserved amongst species, and it encodes a DNA-binding protein that is believed to play a role in RNA splicing mechanisms (E. Kim et al., 2014). It modulates β-secretase levels and *APP* cleavage (C. Zhang et al., 2010).
- *ADAM10*: α-secretase is encoded by this gene, which cleaves the *APP* protein. Two known mutations hinder this protein's ability, leading to the formation of amyloid plaques. *ADAM10* is also a very likely candidate for the development of LOAD (Tanzi, 2012).
- *BIN1*: A highly expressed protein in CNS, *BIN1* is believed to affect Aβ production and clearance (Dourlen et al., 2019). More importantly, among the genes identified with GWAS, it is the first gene related to forming NFTs. It is believed that functional impairment due to this gene leads to increased Tau loads. BIN1 interacts with Tau. It is shown in mouse studies that overexpression of this gene reduces long-term memory loss (Sartori et al., 2018).

1.2.3. Brain Cell Types and Their Characteristics in AD

Brain tissue consists of two main cell types: Neurons and glial cells. Neurons are cells with many subtypes which communicate through excitatory signals. In contrast, neuroglial cells are the ones that support the integrity of neural communication and protect the structure of neural links through multiple ways, such as yielding structural support for the neurons, repairing damage, myelination, providing nutrition, maintaining the blood-brain barrier and extracellular matrix, and protecting the brain cells from foreign matters (Purves et al., 2012).

1.2.3.1 Neurons

Neurons are the excitatory cells that carry signals through polarization, meaning they have multiple states switching through signal carriage. In their neutral state, they are highly polarized. Through a change of ion concentration between the inside and outside of the cell membrane, these cells locally change state (depolarization), and the signal is carried through the cell body as a wave. As the signal flows, depolarized regions are repolarized, turning the cell back into its neutral and polarized state.

Neurons consist of a perikaryon (body of the cell), dendrites (emerges from the perikaryon), and a single axon, which usually extends much more prolonged than dendrites. Dendrites are the building blocks of grey matter, while axons are the building blocks of white matter. The shapes of dendrites and axons are used in the determination of morphological neuron type. Generally, the signal is carried through the neurons from the dendrites to the perikaryon and the axon, respectively. In that case, the dendrites can be considered as the receivers of the neuron, which expand the cell's surface by reaching out from the perikaryon. The number of dendrites and the branching structure varies depending on the neuron type. For example, a pyramidal neuron with the three-dimensional expenditure of dendrites will receive signals from different cortical layers, while a neuron with dendrites on only a single cortical layer will receive signals only from that layer.

Perikaryon is the part where organelles take place. It is very rich in neurofilaments and microtubules. The axon transmits neural information from the cell it stems from. Generally, due to having a higher number of neurofilaments, it is distinguishable from the dendrites by length and thickness (Purves et al., 2012; Squire et al., 2012).

During a confirmed AD pathology, perikaryon contains NFTs, and the extracellular regions, primarily the synaptic regions, contain amyloid fibrils. These bodies hinder the cell's ability to maintain necessary metabolic processes for its survival and synaptic communication. In addition, surrounding glial cells are activated and start an inflammation in the surrounding area, further interfering with the beforementioned pathways. Ultimately, cell death occurs (Bandyopadhyay, 2021).

1.2.3.2 Astrocytes

Astrocytes, also known as astrocytic glial cells, are the regulators of synaptic signal transmission. They also control the blood volume, nutrition delivery, and concentration of compounds such as ions within the extracellular matrix. They can be considered the housekeepers of the brain tissue. In addition, the factors they produce induce the differentiation of the stem cells into neurons (Fields et al., 2014).

Astrocytes can be found in both the brain and spinal cord. They are the most common cell type found in the human brain, with different sub-types and spatial organizations depending on their localizations. Compared to other animals, they are also highly specialized in human brains (Fields et al., 2014; Freeman & Rowitch, 2013).

A single astrocyte can be in contact with millions of synapses. Since they regulate synaptic functions, they are responsible for synapse formation and plasticity by secreting gliotransmitters such as glutamate and ATP (Agulhon et al., 2010). Having this amount of contact raises the idea that they have a crucial role in intellectual development, meaning they are also responsible for producing new information (Fields et al., 2014). The underlying mechanisms of synaptic plasticity can change depending on the brain region (Agulhon et al., 2010).

As it is their role to repair damage and protect neurons from harm, it can be considered they have two states: A passive state where the cell upkeeps the continuous tasks such as nutrition delivery, and a reactive state where the cell actively repairs neural damage while having upregulated protein synthesis and degradation reactions (Monterey et al., 2021), and increased secretion of growth factors and gliotransmitter to increase the synaptic formation and axonogenesis in addition to secreting cytokines and chemokines (Li et al., 2019).

Reactive astrocytes are separated into two branches: A1 astrocytes and A2 astrocytes. A2 astrocytes are healthy reactive astrocytes that repair damaged neurons. A1 astrocytes are quite the opposite; they have upregulation of genes that are destructive to synapses. Secreted factors from A1 astrocytes have inflammatory effects that damage the blood-brain barrier. Some of these factors can increase

secretase activity that can cleave APP, which leads to amyloid fibrillation (Liaoi et al., 2004). Reactive A1 astrocytes are generated when microglia encounter amyloid plaques and secrete inflammatory factors (Burda & Sofroniew, 2014).

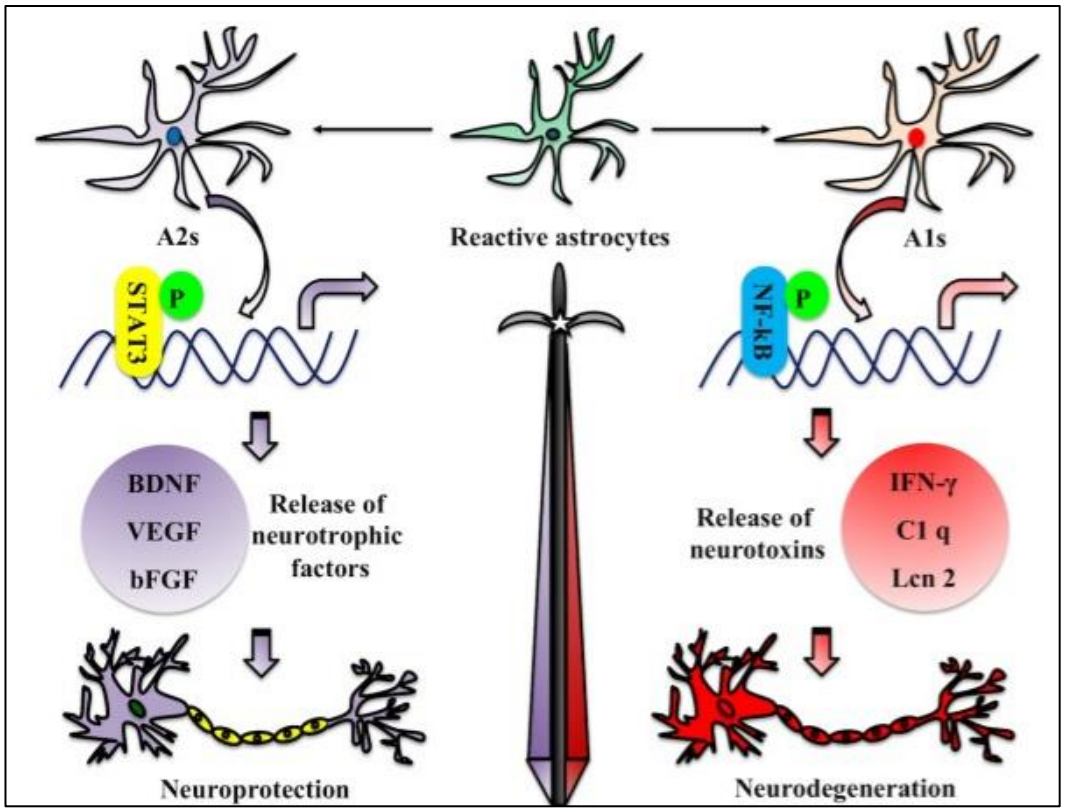

Figure 8 – Reactive astrocytes. A1s induce neurotoxicity, while A2s induce neuroregeneration and protection (Li et al., 2019)

1.2.3.3 Microglia

Microglia are the cells carrying the role of macrophages in the brain and the spinal cord. They are responsible for removing cell debris and unneeded cells or synapses through phagocytosis, and they are the first cells to react to foreign objects found in CNS (Dissing-Olesen et al., 2007). If astrocytes are considered the key players in brain homeostasis, microglia can be the key to maintenance. (dos Santos et al., 2020; Filiano et al., 2015; Ginhoux et al., 2013).

When there are no threats in the tissue, microglia stay in their so-called “resting” state even though they are actively surveying their environment. Their secretions and phagocytosis ability are reduced compared to their active state. If they identify a threat, or encounter cell-necrosis factors, they immediately transform into their active state to remove it by not letting any inflammation occur (Verkhratsky & Butt, 2013). They are also very sensitive to changes in potassium concentration. Since brain tissue cells contain a high amount of potassium, when cell death occurs potassium gradient in the extracellular matrix changes, this alone can switch microglia into their reactive state (Gehrmann et al., 1995).

In regions where AD pathology is observed, microglia are abundant, which indicates AD proliferation is related to microglia (Marlatt et al., 2014). As they are very sensitive to foreign bodies, they activate in the presence of Aβ 1-42 (Guan et al., 2022). This activation proliferates them and causes them to secrete specific cytokines, leading to inflammation. When they are in their active phagocytic form, they clear the amyloid plaques with phagocytosis by surrounding the area and forming a protective barrier (Hansen et al., 2018).

Identification of the plaques is related to the TREM2 receptors on their membrane (Hansen et al., 2018), as TREM2-deficient microglia show reduced clearance. It is also observed that microglial phagocytosis occurs when LDL, *APOE*, and *APOJ* are in a complex with Aβ plaques. (Terwel et al., 2011).

In healthy tissue, these cells can switch between their normal and reactive state, but with Aβ being present, they stay in their reactive form, leading to the formation of a neurodegenerative environment. During the development and adult stages of the brain, microglia are abundant around the synapses, meaning that they can reshape or remove them. Although it is an advantage to remove synaptic connections, it becomes a disadvantage in the presence of Aβ plaques. Their extended reactivation and the environment they create lead to the destruction of synapses and neuron death (Hansen et al., 2018). The inflammation also increases the hyperphosphorylation of Tau proteins (Guan et al., 2022). According to some studies, synaptic loss reduces when microglia are depleted in mice brains with AD (Hansen et al., 2018).

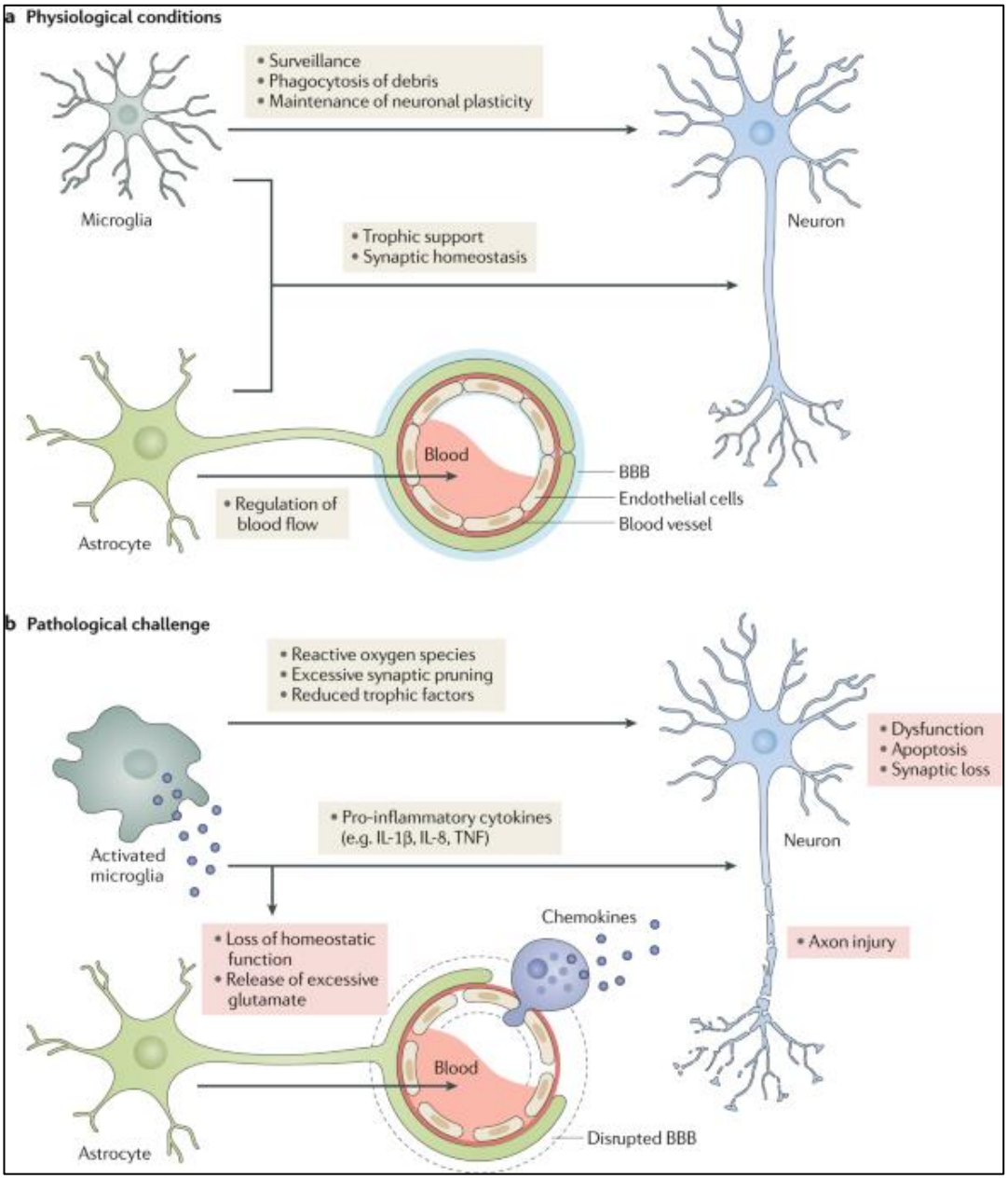

Figure 9 - Astrocyte and microglia in their normal and reactive stages (Leng & Edison, 2020).

1.2.3.4 Oligodendrocytes

Oligodendrocytes are the cells that carry a support role through the myelination of the axons of neurons in the CNS. They extend their membrane and cover a part of the axon body; after that, one side of the membrane elongates and wraps the axon body inwards. This creates a sheath with multiple membrane layers (Richardson et al., 2006). A single oligodendrocyte can operate myelin sheaths on up to 50 axons (Baumann & Pham-Dinh, 2001).

Myelin sheaths increase the insulation of the axon membrane, leading to increased transfer of action potential. According to studies made on rats, the amount of

myelin-sheathed axons is thought to be correlated with intelligence (Juraska & Kopcik, 1988).

As myelin sheaths are responsible for neural transmission, they play a role in cognitive function. The inflammatory environment stemming from Aβ plaques and NFTs is also the reason for damaged oligodendrocytes and, ultimately, myelin sheaths. The inflammation also damages oligodendrocyte precursor cells (OPCs), the cells that generate oligodendrocytes, leading to further demyelination (Cai & Xiao, 2016).

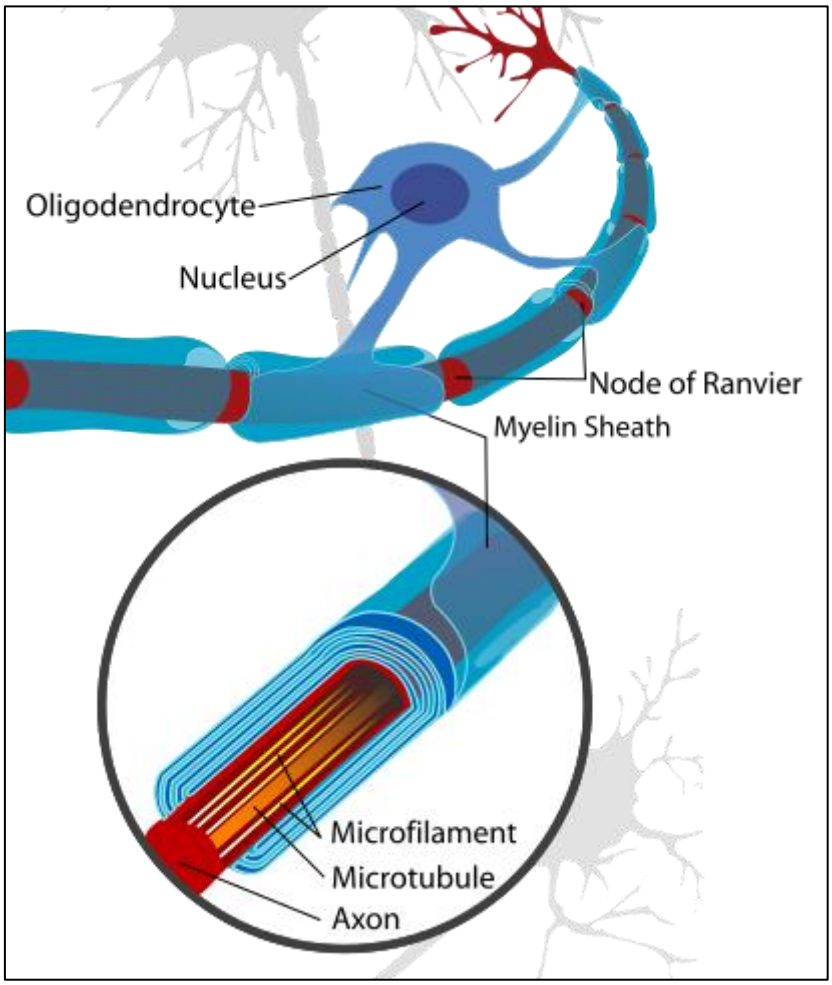

Figure 10 – Oligodendrocyte forming myelin sheath over axons (File:Neuron with Oligodendrocyte and Myelin Sheath.Svg - Wikimedia Commons, n.d.).

1.2.3.5 Ependymal Cells

Ependymal cells are responsible for CNS homeostasis through multiple functions, such as clearing waste and keeping cerebrospinal fluid (CSF) contents under control, by working with astrocytes as they attach to them through their basal membranes. They form ependyma, a structure that separates brain and spinal cord ventricles (filled with CSF) from the neural and glial cells (MacDonald et al., 2021; Staugaitis, 2011).

According to a study, these cells could transform into astrocytes and neuroblasts if necessary, such as when there is a stroke (Carlén et al., 2009).

1.2.3.6 Radial Glia

Radial glial cells provide a structure to newly generated neurons, helping them migrate from the ventricles to the outer layers of the brain (Campbell & Götz, 2002). After neurogenesis that naturally ends during the development stage of a vertebrate, radial glial cells start generating astrocytes (McDermott et al., 2005). They also form ependymal cells on the ventricle surfaces, starting from the postnatal days (MacDonald et al., 2021).

1.2.4. Prevention and Treatment

Due to the difficulty in diagnosis, treatment and recovery of AD are unlikely. The symptoms are observed in the very late stages of the disease, where the hippocampus is disconnected, the network between the neurons is disrupted, and a large amount of cell death has occurred (Allen et al., 2007; Brier et al., 2012). Even if the causes of the disease are lifted at this stage, healing the tissue and restoring the conditions to how it was is very unlikely. Prevention is the best way to approach this disease when these are considered.

A lot of environmental factors are related to the development of AD. These include substance usage, exercise levels, stress, diet, and other diseases such as hypertension or diabetes.

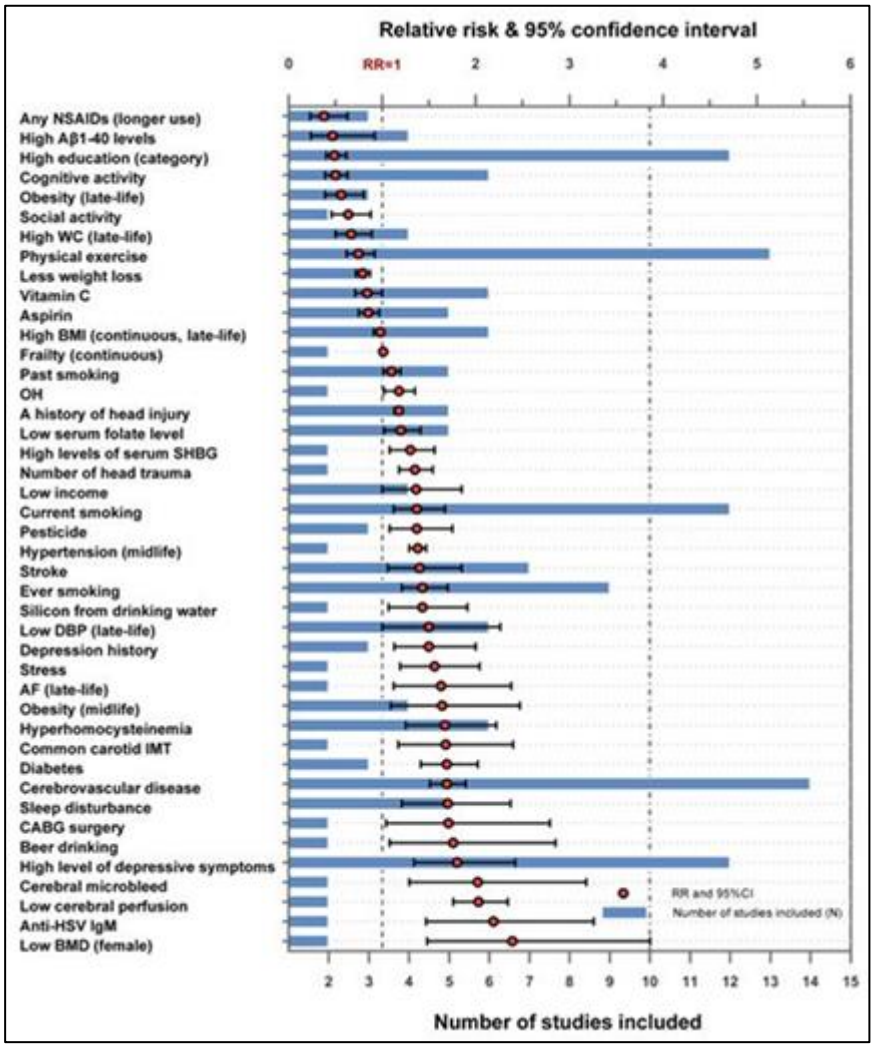

Figure 11 – Risk factors for AD (Yu et al., 2020)

Physical exercises, especially aerobic ones, are shown to increase neuroplasticity and reduce dementia. Mental exercises (such as puzzle solving) and higher education have similar benefits (Cheng, 2016).

Currently, five substances are used in treating the cognitive impairments of people with AD: tacrine, rivastigmine, galantamine, donepezil, and memantine (Fink et al., 2020). These provide a minimal amount of benefit.

Another drug called lecanemab was approved by FDA in 2023. The drug is for AD patients with mild cognitive impairment. It is an IgG1 monoclonal antibody that binds to Aβ amyloids, believed to reduce amyloid markers and lead to the decline of cognitive impairment (van Dyck et al., 2022).

1.3. Sequencing Technologies

"Sequencing" can be defined as solving the order of nucleotides of the genomic information carrier molecules, mainly the DNA.

With the invention of polymerase chain reaction (PCR) in 1983, it became possible to amplify a region of interest of DNA. This has led to the development of many

techniques for analyzing DNA fragments. One of these techniques was the first generation of sequencing. In the following years, more advanced techniques for sequencing were developed. It should be noted that all of these methods are based on the PCR principles.

The PCR technique requires two primers targeting a region on the DNA, free nucleotides (dNTP, deoxynucleotide triphosphate), which are the building blocks of new chains to be synthesized polymerase enzyme that is stable in high temperatures, and a mixture of compounds that can stabilize the reaction.

The primers are carefully designed to bind to the terminus of the targeted region. This interaction must be specific to avoid amplifying other regions in the DNA. They should not form hairpin structures (folding and interacting with themselves) or create primer dimers (interaction of the primers).

The polymerase needs to withstand high temperatures since, in some steps of PCR, DNA chains have to be separated (denatured) by the introduction of heat to the mix. Depending on the experiment's sensitivity, polymerases with proofreading (replacing the mismatched nucleotide) activity are preferred.

In addition to the nucleotides, cations such as Mg^{+2} are added to the final mix. This can stabilize the enzyme or act as a co-factor (Carr & Moore, 2012).

The PCR procedure consists of 3 main steps: 1) Denaturation, 2) Annealing, 3) Extension. The denaturation step introduces the highest amount of heat to separate the DNA chains. When the annealing step starts, the heat drops to the melting point (known as Tm), the temperature at which primers bind to the DNA. Lastly, the temperature is increased again to the point where the polymerase works optimally, and starting from where the primers are bound, the extension of the second chain begins (Sharkey et al., 1994).

These three steps are repeated as needed. In every cycle, the amount of the targeted region increases geometrically.

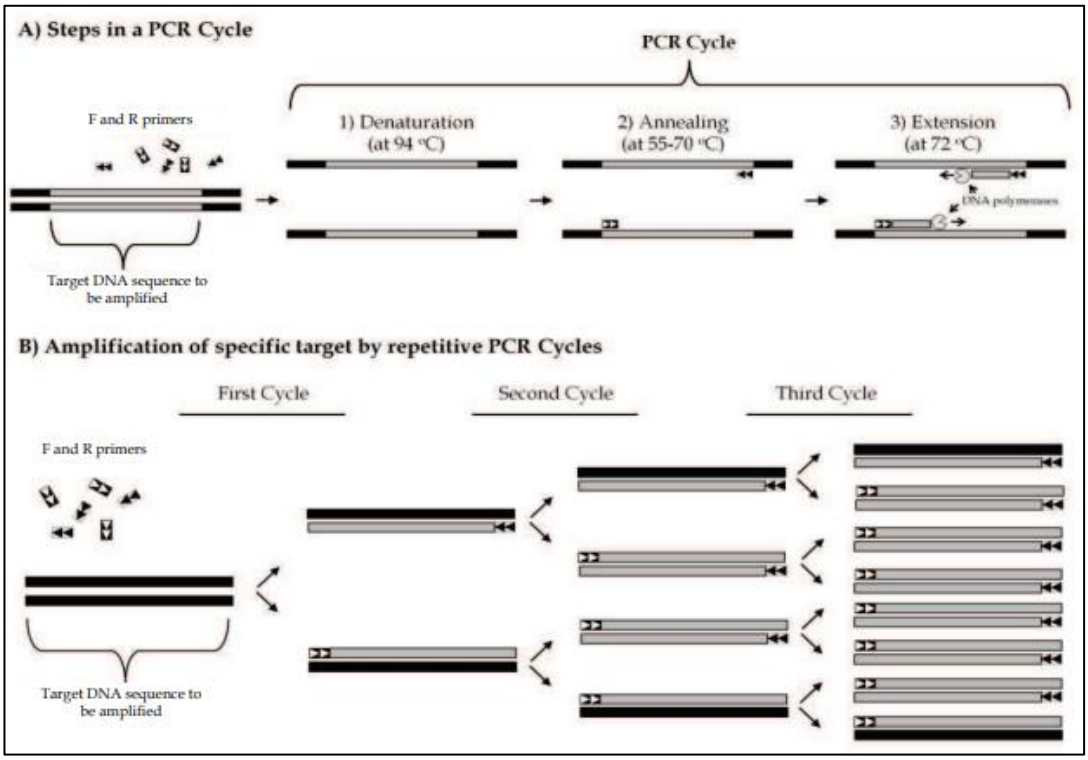

Figure 12 – PCR steps. A) Steps of a PCR procedure B) Geometric amplification of target DNA with PCR (Yilmaz et al., 2012).

1.3.1. First-Generation Sequencing

First-generation sequencing, also known as Sanger sequencing, was developed by Frederick Sanger and colleagues in 1977. The method uses di-deoxynucleotide triphosphates (ddNTP), which stops the elongation process if added to the chain because they do not have the 3'-OH group that is needed to form the phosphodiester bonds.

The method allows for the sequence to be read by having four different PCR reactions. Every reaction contains all types of standard dNTPs, but only one kind of ddNTP is added to each of them, i.e., the tube containing ddATP will have long chains with different lengths that only have the "A" nucleotide at the 3' terminus. The same applies to the reactions with other ddNTP molecules.

These four reactions are added to 4 different lanes in a gel that can let DNA fragments pass through in an electrical field. The shorter fragments can travel a longer distance. Every amplified fragment will form a visible band on the gel. Since the relative length of the fragments and the 3' terminus nucleotide is known in each lane, these bands can be read as letters representing the nucleotides from bottom to top, revealing the sequence of the targeted DNA region.

With the introduction of fluorescent dyes, the sequencing process can be done in a single lane, with each 3' nucleotide emitting a different signal that is processed as color (Sanger et al., 1977).

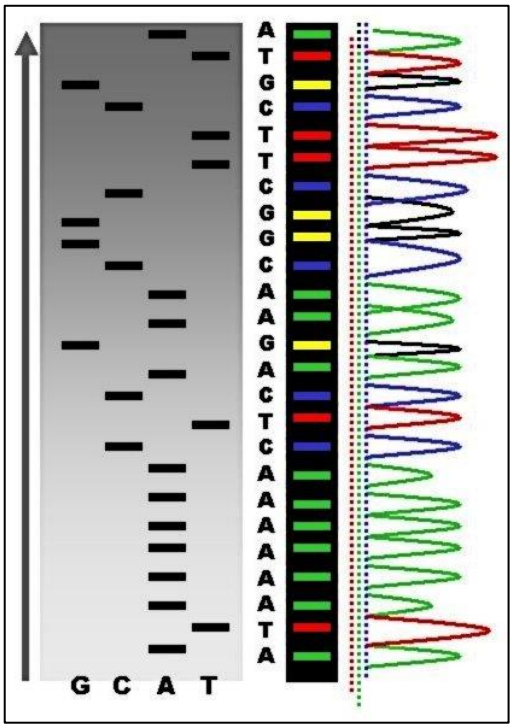

Figure 13 – Resulting bands from a Sanger sequencing experiment. The shorter fragments are at the bottom. The sequence is read from the bottom to the top. With the addition of fluorescent signaling, it is possible to separate the bands ending with different dNTPs in a single well (Wikipedia, 2022).

1.3.2. Next-Generation Sequencing

Next-generation sequencing (NGS), also known as second-generation sequencing or massively parallel sequencing, is a set of different methods that depends on the fragmentation and the amplification of the target region. All of the methods under this branch have similarities in terms of generating large amounts of sequence data.

These methods also rely on synthesizing the second strand of the DNA. The initial step for library preparation is the random fragmentation of the targeted region(s) or whole genome. These fragments are amplified by using random primers (random priming). Adapter ligation also occurs in this step, which allows the fragments to be attached to a physical where the denaturation and second-strand synthesis occurs, which can happen simultaneously for millions of reads. Every time a dNTP is added to the strand that is being elongated, a signal is received (generation of this signal depends on the method that is used), which contains the information about the added dNTP. Since this happens in parallel with many other fragments, a large amount of data can be generated quickly compared to Sanger sequencing (Shendure & Ji, 2008).

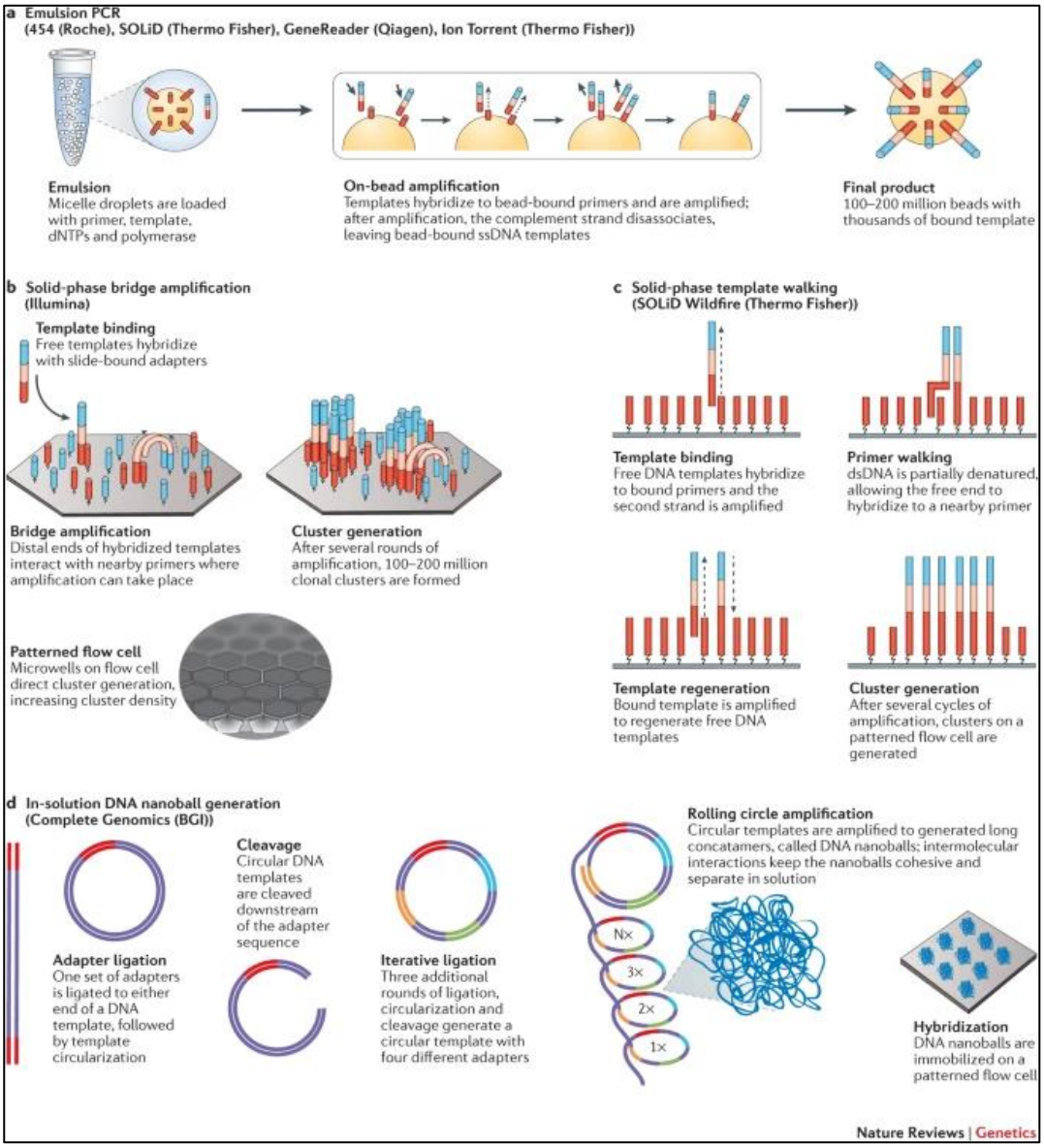

Figure 14 – Current NGS techniques. (Goodwin et al., 2016)

1.3.2.1 Analysis of Sequencing Data

Gathered information from a sequencing experiment is generally stored in two types of file formats: FASTA and FASTQ. FASTA files contain a header for each read that identifies the read by assigning it a name, ID, and other information that may contain details about the experiment. In FASTQ files, however, in addition to the information line, like in FASTA format, there is another line that keeps the quality score of each nucleotide. During the sequencing process of the whole experiment, the machine reads the signals coming from a single point. If all the signals represent the same nucleotide, meaning that they are in the same color, the machine makes a guess with 100% certainty about the type of that nucleotide. The level of certainty is transformed into a human-readable format called PHRED score, or in other terms, quality. In the real world, 100% certainty is unachievable. Due to the limitations of the experiments and errors introduced, the signal may vary, and the certainty

drops. If the converted value of this certainty (the quality score) is below a certain threshold, which is considered 20 as a rule of thumb, the read from that position is considered low quality.

```
1   @ERR000589.41 EAS139_45:5:1:2:111/1
2   CTTTCCTCCCTGCTTTCCTGGCCCCACCATTTCCAGGGAACATCTTGTCAT
3   +
4   3IIIIIIIIIIIII>1IIIFF9BG08E00I%IG+&?(4)%00646.C1#&(
5   @ERR000589.42 EAS139_45:5:1:2:1293/1
6   AGTTGTTAAAATCCAAGCCAATTAAGATAGTCTTATCTTTTTAAAAGAAAT
7   +
8   IIIIIGII.AIIII=?I9G-/II=+I=4?761BA2C9I+5A711+&>1$/I
```

Figure 15 – Example of a FastQ format. There are 8 lines, meaning that there are 2 reads. For the first 4 lines, the read can be seen in line 2, and the information regarding the read is positioned in line 1. Line 3 only contains a '+' symbol which doesn't carry any information. Line 4 contains the quality scores for each position of the read (FASTQ File Format Example., n.d.).

The quality tends to drop as the reads get closer to the 3' end. This is due to the loss of synchronization in the second strand synthesis in a cluster of fragments. The drop in the quality towards the ends of the reads is considered normal.

The quality may also drop from certain defects in the grid that holds the fragments. Contaminations can be another reason for the drop in quality. Usually, the source of the contamination is the adapter sequences used in the experiment.

The bases with low qualities at the ends of sequences and the adapters found within the reads are removed by tools specifically built for these tasks. This operation is termed "trimming." Sometimes, the quality drops to such a level that the machine cannot decide which nucleotide resides at the given position, so that position is assigned with the letter "N." If these are abundant, another operation called 'masking' must be applied that removes these positions with 'N' from the reads. These tools can also practice adapter removal (Anderson & Schrijver, 2010; Ballard et al., 2020).

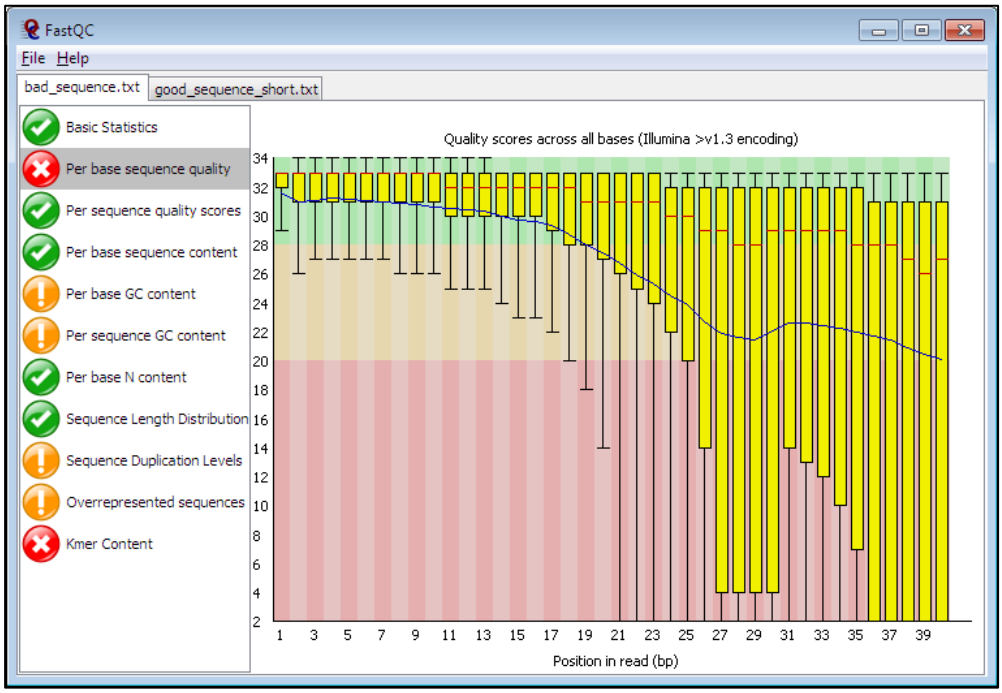

Figure 16 – An example analysis of a FastQ file with FastQC. The figure shows the average quality for each position considering all reads in the file. It can be seen that as we move towards the ends of the reads, the quality drops drastically. This example graph tells us we need to trim the low-quality bases from the 3' ends of the reads (Babraham Bioinformatics - FastQC A Quality Control Tool for High Throughput Sequence Data, n.d.).

The reads are analyzed with tools such as FastQC for their qualities, GC contents, length distribution, and the level of the overrepresented sequences within the read file. For genomic sequencing projects, the overall abundance of each read is expected to be close to each other. If a single sequence or a set of sequences is highly abundant, it is probably because of a PCR bias or contamination. If necessary, these overrepresented sequences can also be removed through a step called 'deduplication.' In experiments like RNA-Seq, seeing overrepresented sequences is expected because, at a given moment, a cell might have overly expressed genes relative to other genes. Seeing this expression profile represented in the read counts is expected.

After analyzing the reads with a tool like FastQC and going through trimming, masking, or adapter removal steps if necessary, the reads are analyzed once more to see if any problems are introduced through the cleaning process. If everything looks as expected, the alignment process begins.

```
cutadapt -a AACCGGTT -o output.fastq input.fastq
```

Figure 17 - An example adapter removal process command with cutadapt. Here, the adapter 'AACCGGTT' is to be removed from the reads in 'input.fastq', and the resulting file will be written into the same directory as 'output.fastq'.

Usually, in the alignment process, if the reads are coming from a known organism, there is a reference sequence for that organism, and it is used as a map for the

reads, and the reads are aligned to the positions that they match within the reference genome. If the read qualities are high, the reads can map to the genome with a high mapping quality. It is expected for multiple reads to align to the same position so the observer can be certain that this is not a random alignment. The number of reads aligned to the same region represents that location's 'alignment depth'. Every read added to a single region is counted as an 'X.' As a rule of thumb, a minimum of 30X is considered a good amount of depth. The reads that consecutively align next to each other cover the genome, termed 'coverage.' For example, whole exome sequencing (WES) experiments will have low genome coverage as the reads are coming from the exonic regions. On the other hand, whole genome sequencing (WGS) will have very high coverage of the total genome. If a sequencing experiment targets only a specific region and the DNA samples are prepared for this kind of sequencing, the total genome coverage will be very low, while the depth for the targeted regions will be very high.

The reads might have unmatched regions in the position they are aligned to the reference genome. Considering we have a high amount of mapping quality and depth, if we see a different nucleotide present in the same position on the genome for all aligned reads, we can say that the sample that provided the reads carries a variant for the location. If half of the reads carry this variant, it can be considered a heterozygous variant, while if all of the reads have the variant, it is a homozygous variant, with exceptions. In addition to mismatches, it is possible to find insertions or deletions within the aligned reads.

The alignment process requires a genomic read file containing an organism's reference genome. These are usually in FASTA format. The reads coming from the FASTQ files are aligned to the reference genome with tools depending on the type of the reads (genomic, exonic, etc.). While DNA reads can be directly aligned, reads from mRNAs might be from different splice variants, meaning that intronic regions must be considered while aligning. Tools like RNA-STAR or HISAT2 consider this and are used for RNA-Seq alignments. The alignment process produces a file in SAM (Sequence Alignment Map) format, a text-based format that contains positional information for every read that was tried to be aligned, meaning that this file contains the contents of FASTQ file(s) and the FASTA file that contains the reference genome, making the file size larger than the sum of all of the inputs.

Since text-based storing is very inefficient as a digital format, the SAM files are converted into BAM (Binary Alignment Map) files in binary format, hence the name. Depending on the compression amount, BAM file sizes can be drastically small compared to their SAM counterparts. The conversion of SAM files can be made with tools such as SAMTools. These files can also be analyzed by coverage, depth, etc.

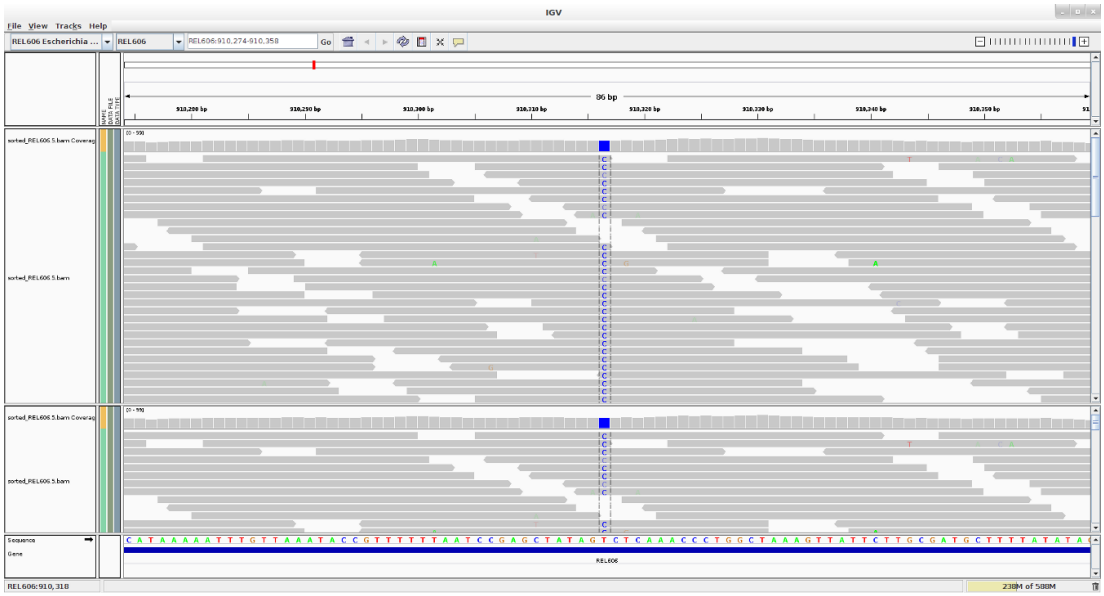

Figure 18 – An example of two BAM files shown in IGV. Every gray bar represents an aligned read to the reference genome, which is positioned at the bottom of the window. Blue colored 'C' represents a variant, and it can be seen in all of the reads (Integrative Genomics Viewer (IGV) Tutorial - Bioinformatics Team (BioITeam) at the University of Texas - UT Austin Wikis, n.d.).

The aligned reads can be visualized by tools such as Integrated Genomics Viewer (IGV). These tools work with the BAM format.

Depending on the initial data type, the downstream steps can change a lot. For WGS or WES, the analyses are made to find the variants among the samples. For exonic files, an expression profile is created for each sample. The distribution of the variants in a given position amongst the reads is considered, and tools that test the statistical significance of these variants are used to generate a format that keeps the positional and numerical information of them (Variant Call Format: VCF). In RNA-Sequencing, an expression profile is created for each sample; a count matrix is generated that contains the reads aligned to each gene, providing the amount of expression for them. The samples are grouped depending on their conditions, and the amount of expression for each gene is compared in between to understand if a particular condition creates a difference in the expression profile. Genes with a high amount of difference in expression are considered significant for the given condition. While generating the VCF file or the count matrices, an 'annotation file' is given as the second input along with the BAM files. These annotation files keep the positional information (chromosome, position on the chromosome, strand, length, ID) about genomic elements such as genes, exons, UTRs, pseudogenes, promoters, and enhancers, and start and stop codons.

Both analysis methods take important functional regions (genes, exons, etc.) as a list. If needed, these regions are provided to certain tools which search for common molecular functions, metabolic pathways, diseases, cell structures, etc., to make sense of the lists of genomic elements gathered from previous analyses.

1.3.2.2 Problems with Sequencing Data

A next-generation sequencing experiment has multiple steps that can introduce errors to the final results.

During the library preparation, a PCR step is used to increase the number of fragments to be sequenced. This is necessary because as the number of fragments increase, there will be a higher chance for the reads to be ligated with adapters and bound to the sequencing grid. If the number of fragments bound to the grid is high, the signal strength and, ultimately, the certainty of the nucleotide type will increase, leading to FASTQ files with higher average-quality of reads. Yet, the random primers used in this first amplification step with PCR might introduce biases. As the reads are increasing geometrically in every step of PCR, some fragments will have a lower chance of duplicating. This can lead to loss of information. This bias can be seen in the 'overrepresented sequences' module of FastQC and can be dealt with by using the 'deduplication' method. To avoid this issue from the start, another method is to create technical replicates: Multiple sequencing experiments are done on the same sample.

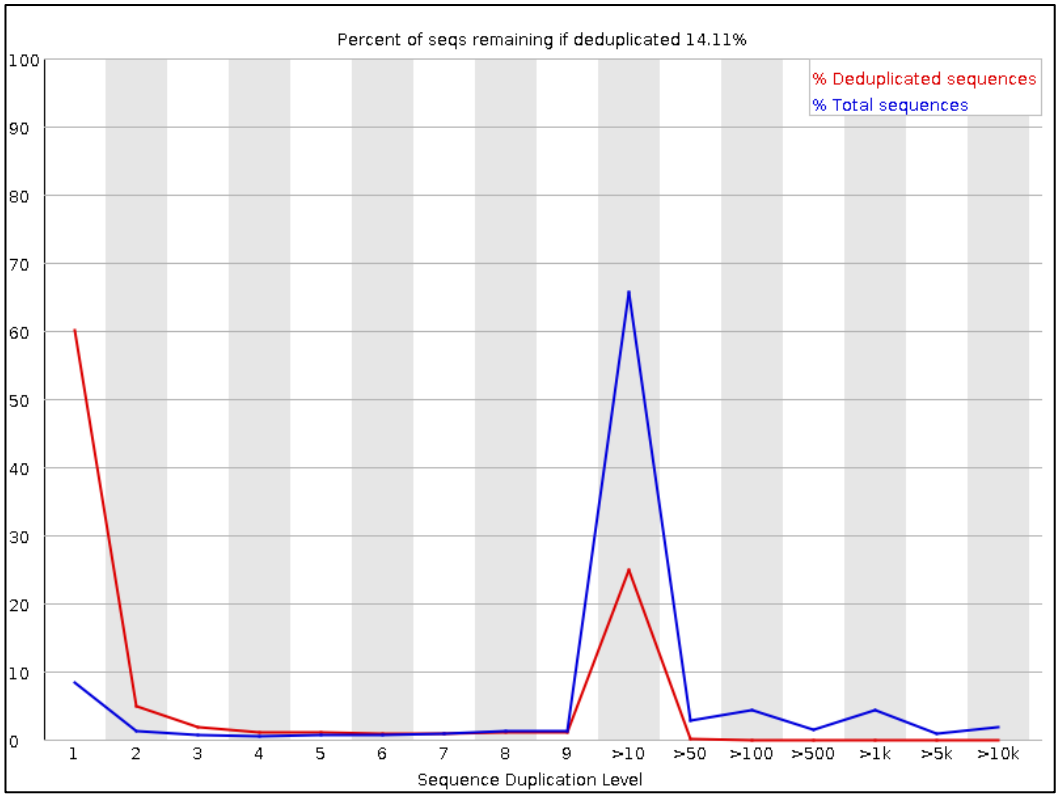

Figure 19 – An example graph from the "Sequence Duplication Levels" module of FastQC. We can see that around 60% of the reads in the file are duplicated more than ten times (How to Interpret This Duplicate Sequence Plot from FastQC, n.d.).

Another problem is the error rate of the enzyme that is used for PCR. Every DNA polymerase has an error rate, which can create a mismatch during the elongation step. If a PCR bias occurs for the strand with the mismatch, it will create a false positive when looking for a variant for that position. There is no downstream

method to avoid this problem. This can again be avoided by creating technical replicates (Ballard et al., 2020).

The most significant limitation of next-generation sequencing methods is related to the repeating regions within the genome. These experiments cannot be done for long sequences. Although the length can vary depending on the method that is used, the maximum length is around 200 base pairs. The fragments acquired from a repeating region will create reads which can align to multiple positions for the given region. Unfortunately, the only way to solve this is to use longer fragments for sequencing, which is not possible as the length of the fragment increases, the uncertainty for the 3' terminus nucleotides also increases. Paired-end sequencing (where the fragment is read from both termini) can overcome this to a scale, but it is still not a solution. If high throughput data is not needed, Sanger Sequencing can be used as it produces read length with 400 to 900 base pairs. Third-generation sequencing methods were suggested to overcome the short-read problem.

1.3.3. Third-Generation Sequencing

Third-generation sequencing is a relatively new method of sequencing that is specifically built to deal with the short-read length sequencing limitation of second-generation sequencing. This method is also referred to as long-read sequencing.

Third-generation sequencing has not replaced NGS technologies, as they are still under development and have a very high error rate during sequencing.

Currently, there are two widely known methods used in third-generation sequencing: 1) Single Molecule Real-Time (SMRT) sequencing from Pacific Biosciences (PacBio) (Eid et al., 2009), 2) Nanopore sequencing from Oxford Nanopore Techniques (ONT) (Branton et al., 2008).

In SMRT sequencing, DNA (or RNA) taken from the sample is ligated with adapters. A single adapter connects to the 3' terminus of one strand and the 5' terminus of the opposite strand. This is done for both termini of the double-stranded DNA (dsDNA). With these adapters' ligation, a circular single-stranded DNA (ssDNA) is obtained when the dsDNA is denatured. DNA polymerase is attached to this DNA, and the polymerase-DNA complex is directed to a well which is called "zero-mode waveguide" (ZMW) on a surface. There are dNTPs with distinct fluorescent tags within these wells. As the second strand synthesis begins, the fluorescent tag is removed by the reaction when a dNTP is attached to the strand, revealing the identity of the dNTP. Both strands of a long DNA fragment can be sequenced this way, with the adapters between them (Eid et al., 2009).

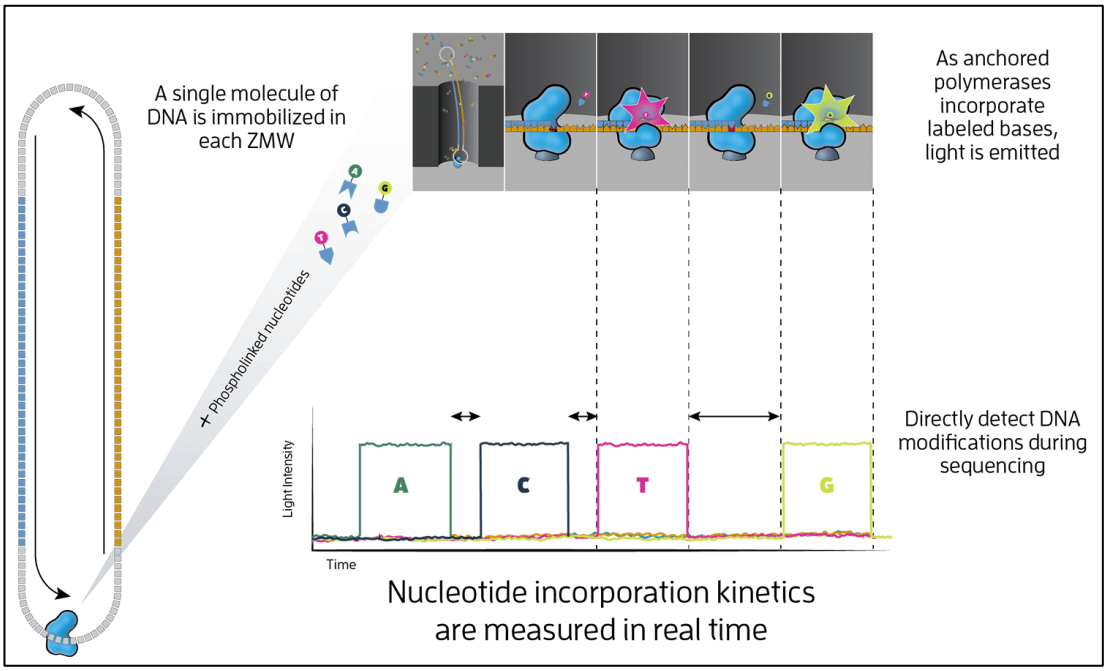

Figure 20 – PacBio SMRT sequencing. On the left, the DNA fragment with both of its chains (blue and orange) is shown. The strands are separated but connected on the termini with adapters (shown in gray). Polymerases are anchored within ZMWs, and a specific fluorescent signal is obtained during the second strand synthesis when a dNTP is bound (How HiFi Sequencing Works - PacBio, n.d.).

ONT uses special porin molecules that form a pore 10nm in diameter. These porins are placed on an electrically resistant membrane while having ions flow through the pore, forming a net electrical charge. When a molecule passes through the pore, this charge is disrupted. The disruption can be measured to identify the molecule.

In the ONT technique, one end of the dsDNA (or RNA) is ligated with an adapter to form a hairpin structure. A special helicase enzyme attaches to the other end and guides the DNA to the pore structure. The helicase connects to the porin and starts separating the DNA strands. One of the strands passes through the pore and disrupts the electrical current, leading to the identification of the nucleotide which passes through. Both strands can be sequenced this way thanks to the hairpin structure of the other end. As opposed to all techniques explained before, ONT sequences the molecule by the helicase activity instead of a polymerase activity (Branton et al., 2008).

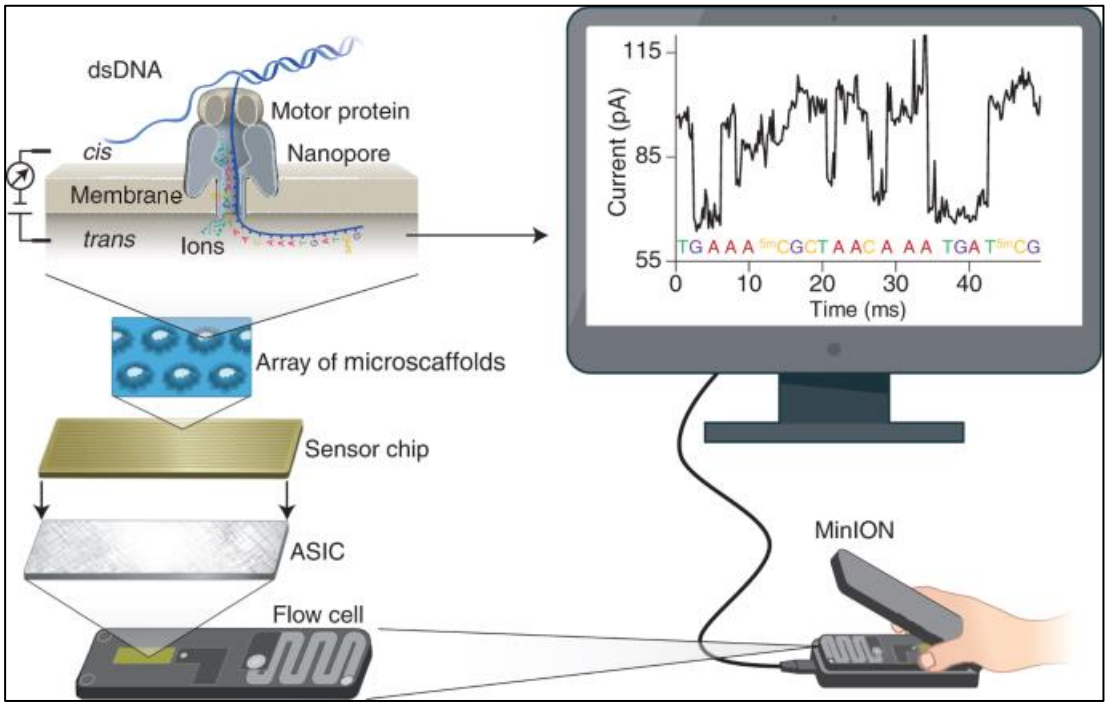

Figure 21 – Oxford Nanopore sequencing. A protein with helicase activity separates the strands of the DNA fragment, and while the strands are passing through the pore, the nucleotide is identified by the disruption it creates on the electric field (Y. Wang et al., 2021).

These two techniques for sequencing don't necessarily require prior fragmentation or amplification with PCR. This way, long reads can be obtained without PCR bias (Bleidorn, 2015; Branton et al., 2008; Eid et al., 2009).

Since these methods are relatively new, they also bring challenges, one of them being their high error rate. In SMRT sequencing, the polymerases are damaged as they continue sequencing (Gupta, 2008). In ONT sequencing, the resolution of sequence reads is low due to the fast-moving speed of the nucleotides through the pore, which creates background noise and makes nucleotide reads hard to generate (Loman & Watson, 2015).

CHAPTER 2

MATERIALS AND METHODS

2.1. Gathering and Manual Filtering of the Raw Data

The raw data for this project was downloaded from the official website of the National Center for Biotechnology Information (NCBI) (Sayers et al., 2022). Instead of directly downloading from the Sequence Read Archive within the NCBI website, the search was performed in the "BioProject" database to find matching "Control vs. Patient" samples. "(Alzheimer) OR (AD)" was used as the search term. The results were further filtered by selecting the data type as "transcriptome," scope as "multi-isolate," and organism as "human." The filtered data does not contain Gene Expression Omnibus (GEO) datasets because the raw data will be processed through the same methods and tools for each sample, and GEO datasets could have different methods than what is used in our workflow, which has the potential to affect the results.

The data was then re-filtered by manually checking the experiment design and samples. Data without mRNA reads (such as microRNA studies), samples without a clear definition of control vs. diseased, samples with only control vs. treatment, and scRNAseq data were excluded. Although it was possible to perform pseudo-RNA-seq on scRNAseq data, they were still excluded due to their size and computational requirements. They could also create a bias caused by the experimental differences they introduce to RNASeq studies.

After the filtering process, out of 94 total projects, 23 projects were selected for analysis. Within these 23 projects, 4 of them did not have precise categorization for their samples and were also excluded. In the final list, there were 19 projects to take to downstream analysis (See Appendix A for further details about the samples and the experimental designs).

The selected samples are PRJNA232669, PRJNA279526, PRJNA399530, PRJNA451437, PRJNA482601, PRJNA527202, PRJNA559812, PRJNA576835, PRJNA603192, PRJNA643561, PRJNA644383, PRJNA662330, PRJNA675864, PRJNA683625, PRJNA688060, PRJNA688885, PRJNA714081, PRJNA727602, and PRJNA767074 (Cardona et al., 2021; Chen et al., 2021b; Gao et al., 2022; Garofalo et al., 2020; Hanna et al., 2021; Ivashko-Pachima et al., 2021; Konttinen et al., 2019; Kwart et al., 2019; Lefterov et al., 2019; Lin et al., 2018; T. Liu et al., 2020b; Magistri et al., 2015; McQuade et al., 2020; Meyer et al., 2019; Mizuno et al., 2021; Nativio et al., 2020; Oksanen et al., 2020; Potemkin et al., 2022; Scheckel et al., 2016).

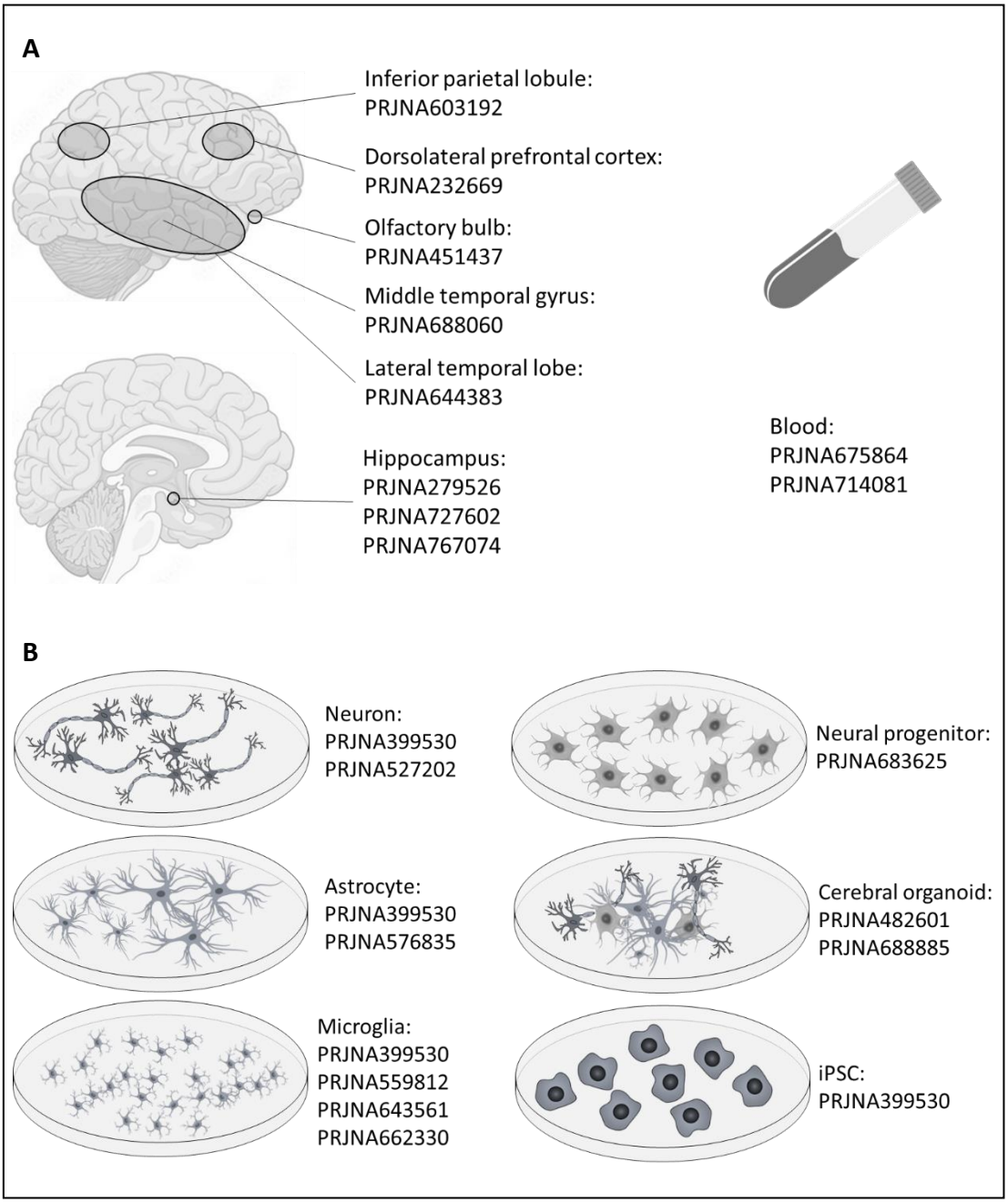

Figure 1 – The list of projects that were used in the downstream analysis. A: Projects with the tissue samples B: Projects with the cell culture samples

2.2. Data Processing

The following steps used Linux Mint 20.3 x64 with GNOME Shell 3.36.9 as the working environment. The environment had access to 24GB of SODIMM 2133Mhz DDR4 RAM with an Intel 6700-HQ CPU. Each of the following tools was installed following instructions on their official websites.

2.2.1. Acquiring the Data

Each run file in each project was downloaded via the "prefetch" tool (version 2.11.2) from "SRAToolKit" provided by NCBI. The downloaded files with .sra format were then validated with "vdb-validate" v2.11.2, again provided by SRAToolKit (*SRA Toolkit* , n.d.). After the validation, files were processed into fastq files using "fasterq-dump" v2.11.3. Instead of "fastq-dump --split-3", fasterq-dump was used as suggested in the manual of SRAToolKit.

2.2.2. Data Analysis and Cleaning

The data quality was analyzed via FastQC v0.11.9 (*FastQC A Quality Control Tool for High Throughput Sequence Data*, n.d.). None of the data required masking. The main problem was adapter contamination and quality drop at the end of the sequences. Overrepresented sequences and per-base sequence content modules of FastQC raised a warning for all files, which is expected in RNA-Seq outputs. Since, at a given moment, a cell might have overexpressed genes, and in the experimental design of RNA-Seq studies, there might be biases at the beginning of the reads; these warnings were ignored.

The trimming was performed with cutadapt v2.8 (Martin, 2011) on each fastq file, regardless of FastQC warnings in the quality modules, with the following parameters:

cutadapt -q 20 -m 20 -a [ADAPTER]

This is to ensure there are no base reads with a quality lower than 20 allowed at the 3' ends of the sequences and no sequences shorter than 20 bases. The '-a' parameter takes a given adapter and removes the adapter and the subsequent bases towards the 3' end. This was performed on single files for single-end studies and on pairs at the same time for paired-end studies.

Trimmed fastq files were re-analyzed via FastQC to ensure no other unexpected consequences occurred from the trimming step. After making sure there were no problems within the files, they were taken into the alignment step.

2.2.3. Acquiring the Reference Genome and Genome Annotations

The reference genome was downloaded from the Ensembl database: "Homo_sapiens.GRCh38.dna.primary_assembly.fa" (Cunningham et al., 2022). Genome annotations were also downloaded from the Ensembl database as "Homo_sapiens.GRCh38.105.gtf". The selected GTF file contains all annotations for the human genome. A filtered version was not selected at this step since the filtering process could be applied later if needed.

2.2.4. Aligning the Reads to the Reference Genome

For RNA-Seq alignment, three tools were considered: RNA-STAR, HISAT2, and Bowtie2. Since the required RAM capacity of RNA-STAR is over 30GBs, it was not used. HISAT2 (version 2.1.10) (D. Kim et al., 2019) was selected for alignment as it is more up-to-date compared to bowtie2.

HISAT2 works with indices of the reference genome, using less than 7GB of RAM for a single alignment process. Following the instructions, .hisatindex files were built from the reference genome. After this, every fastq file was aligned to the reference genome, and the corresponding alignment files (SAM files) were generated. Single-end reads were aligned with the '-U' parameter, and paired-end reads were aligned with the '-1' and '-2' parameters.

As SAM files are inefficient for both storage and downstream steps, they were converted into BAM files via the "view" tool of the "Samtools" v1.10 package (Danecek et al., 2021). The BAM files were sorted and compressed with "samtools sort" with the compression parameter of 9 (-l 9), which is the highest amount of compression. BAM index (bai) files were generated for every BAM file using the "samtools index."

```
prefetch -p SRR                         # provides binary read file; -p for seeing the progress
vdb-validate SRR                        # this is control the integrity of the downloaded files
fasterq-dump -e 8                       # use this many cores
             -p                         # show progress
             SRR                        # converts binary read file into fastq
fastqc SRR_1.fastq                      # analyze the fastq files
fastqc SRR_2.fastq                      # use '-t 2' to use 2 cores; analyze two files at the same time
cutadapt -j 8                           # use this many cores
         -q 20                          # trim bases with quality lower than this (use 20,20 if 5' end also needs a trim)
         -m 20                          # remove reads that are shorter than this
         -a ATACATT                     # remove this adapter from pair 1's 3' end
         -A ATACATT                     # remove this adapter from pair 2's 3' end
         -o SRA_1_trimmed.fastq         # pair 1 output
         -p SRA_2_trimmed.fastq         # pair 2 output
         1> SRA_cutadapt.txt            # keep information about trimming process in a file
hisat2 --summary-file SRR_hisat2.txt    # keep information about alignment process in a file
       -p 8                             # use this many cores
       -x hg38_hisatindex               # these are the index files
       -1 SRA_1_trimmed.fastq           # input forward reads
       -2 SRA_2_trimmed.fastq           # input reverse reads
       -S SRA.sam                       # output file name
samtools view -@ 7                      # use additional 7 cores
              -b                        # convert to BAM
              -o SRA.bam                # output name
              SRA.sam                   # input name
samtools sort -@ 7                      # use additional 7 cores
              -l 9                      # compression level; between 1-9
              -o SRA_sorted.bam         # output name
              SRA.bam                   # input name
samtools index -@ 7                     # use additional 7 cores
               -b SRA_sorted.bam        # input name
               SRA_sorted.bam.bai       # output name
```

Figure 23 – The workflow of BAM generation. These are the steps from downloading the files containing the reads (SRR) until the BAM file generation.

2.2.5. Strandness of the Experiments

Before annotating and counting the aligned reads, the strandness information must be gathered. "inferExperiment" v4.0.0 tool was used to gather this information (L. Wang et al., 2012). This tool requires a BAM file and an annotation file in BED format for input, so the GTF file for the human genome was first converted into the BED format using "bedops" 2.4.40 (Neph et al., 2012). inferExperiment outputs the information of strandness for the reads as 1) unstranded, 2) stranded, and 3) reversely stranded.

2.2.6. Generating Count Matrices for Genes

"featureCounts" v2.0.0 was used to count the reads aligned to each gene. '-s' parameter is set to '0' for unstranded reads, '1' for stranded reads, and '2' for reversely stranded reads (Liao et al., 2014). This tool can take multiple BAM files as input, but it was deemed unnecessary as it produces a lot of unneeded information. Each BAM file was given separately.

In this workflow, featureCounts generates count files for each sample separately. For each project, the gene count columns were merged into a single file, which creates what is known as "the count matrix." Every row corresponds to a gene, while every column corresponds to a sample, and the values show the number of aligned reads to that gene in the sample.

```
infer_experiment.py -i SRA_sorted.bam   # input file
                    -r reference.bed    # reference annotations in BED format.
                    > SRR_inferred.txt  # keep the information in a file
featureCounts -T 8                      # use this many cores
              -p                        # use this, if the reads are paired-end
              -a reference.gtf          # reference annotation in GTF format
              -s 0,1,2                  # 0: unstranded, 1: stranded, 2: reversely-stranded
              -o SRA_counts             # counts file as the output
              SRA_sorted.bam            # input file
              2> SRA_featureCounts.txt  # store the information in a file
```

Figure 24 – The steps for acquiring the strandness information and count matrix generation.

2.3. Differentially Expressed Gene (DEG) Analysis

The count matrices generated from featureCounts were given as input to the "DESeq2" tool (v1.36.0) (Love et al., 2014), a package installed via "BioConductor" (v3.16) within the "R Programming Language" (v4.2.2) (Tuimala & Kallio, 2013). DESeq2 requires a design matrix for each count matrix containing specific information about every sample. The information includes if the sample is taken from a patient or a healthy individual, if they are male or female, etc.

DESeq2 requires the count matrix to be non-normalized, as it handles the normalization process itself. While many sub-tools within DESeq2 can generate different outputs, the final output that is necessary for further analysis contains the following:

- baseMean: The mean expression value of a gene amongst the samples.
- log2FC: The fold change of the gene expression value in a patient versus control on the log2 scale. If the value is negative, the gene is underexpressed in patients. If it is positive, the gene is overexpressed.
- lfcSE: Shows the standard error rate of log2FC values.
- stat: Wald statistic
- pvalue: A value that measures the statistical significance of the observed difference in gene expression value in patients versus control. It is calculated with the Wald test.
- padj: Adjusted p-value – the value that further filters p-values in order to avoid false positives on a larger scale depending on the adjustment method, which is Benjamini and Hochberg method by default in DESeq2.

DESeq2 considers the distribution of counts as negative binomial distribution.

Table 4 – A partial example count matrix. The "Genes" column contains the gene IDs. Each column contains the total read count found for that gene. The values shown below are not normalized.

Genes	Sample 1	Sample 2	Sample 3	Sample 4
ENSG00000186827	7	3	8	15
ENSG00000186891	2	9	2	7
ENSG00000160072	234	454	341	511

ENSG00000260179	3	3	2	5
ENSG00000234396	1	0	0	0
ENSG00000225972	11	18	22	26
ENSG00000224315	36	55	34	33
ENSG00000198744	51	41	56	81
ENSG00000279928	3	2	2	3
ENSG00000228037	7	7	6	9

Table 5 – An example design matrix. Every column provides categorical information about the samples. The "run" column is only needed when there are technical replicates within that sample, and they have to be collapsed. These variables are used in DESeq2 for grouping the samples, and then groups are compared to each other for calculations.

	tissue	condition	run	type	sample
SRR2422918	dorsolateral prefrontal cortex	healthy	SRR2422918	paired-end	H1
SRR2422919	dorsolateral prefrontal cortex	healthy	SRR2422919	paired-end	H2
SRR2422920	dorsolateral prefrontal cortex	healthy	SRR2422920	paired-end	H3
SRR2422926	dorsolateral prefrontal cortex	AD	SRR2422926	paired-end	AD1
SRR2422927	dorsolateral prefrontal cortex	AD	SRR2422927	paired-end	AD2
SRR2422928	dorsolateral prefrontal cortex	AD	SRR2422928	paired-end	AD3

```
1
2  # Loading data from PRJNA232669
3  dat = read.csv('PRJNA232669_count_matrix.csv', header = T, row.names = 1)
4  info = read.csv('PRJNA232669_design.csv', header = T, row.names = 1)
5  # Creating DESeq2 object
6  dds <- DESeqDataSetFromMatrix(countData = dat,
7                                colData = info,
8                                design = ~condition)
9  # Pre-filtering: removing rows with <10 counts
10 keep <- rowSums(counts(dds)) >= 10
11 dds <- dds[keep,]
12 # Setting the factor level
13 dds$condition <- relevel(dds$condition, ref = 'healthy')
14 ddsColl <- collapseReplicates(dds, dds$sample, dds$run, renameCols = F)
15 # Running DESeq
16 ddsDE <- DESeq(ddsColl)
17 # DESeq results
18 res <- results(ddsDE, contrast = c("condition","AD","healthy"), alpha = 0.05)
19
```

Figure 25 – An example DESeq2 workflow. The count and design matrices are added to the environment on rows 3 and 4. These are then used to create the DESeq2 object (dds) by taking the "condition" as the testing variable on row 8. Genes with less than 10 reads are removed from the object. The design condition, which is "AD vs. healthy" in this example, is transformed into a factor, and the reference of the healthy samples are selected as the reference. DESeq2 is run, and the results are pulled from the ddsDE object (row 16) as needed.

After providing the count matrix and the design table, a DESeq object is created. Within this object, which variable to be used for comparison has to be selected. For most of the projects, only the "condition" column is given as the variable for testing, which specifies if the sample is healthy or AD-like. For these projects, the "design" parameter was set as "~condition." For other projects, like the ones with different cell types or with a provided "sex" column, the design parameter was set as "~condition+cell+condition:cell" and "~condition+sex+condition:sex," respectively.

Generating the results file also requires the variables to be specified. From a single DESeq2 object, multiple results files can be generated. So, it is not necessary to run DESeq2 multiple times with different designs.

In each project, the variables were considered before generating the results. Only "AD vs. healthy" results were generated in projects without information other than the condition. In the tissue samples with sex information, the comparisons were made for female and male samples separately, but the analyses were repeated without including the sex variable. In cell culture samples, the "AD vs. healthy" comparison was made for each type of cell, but the combination of these was not made. As a result, the final comparison list contains the following categories:

- AD vs. healthy, for samples without sex variable
- Female AD vs. female healthy
- Male AD vs. male healthy
- AD vs. healthy, combinations of male and female samples
- Cell type AD vs. cell type healthy

Each DESeq2 results file was then filtered according to the p-values. Genes with p-values lower than 0.05 were considered significant and taken into the downstream analyses.

2.4. Gene Set Enrichment by Gene Ontology Terms

Gene ontology (GO) is a method to identify attributes of a gene by its biological process (BP), molecular function (MF), and cellular component (CC) (Harris et al., 2008).The gene lists with their fold change and significance values from DESeq2 were used as the inputs in this step. Each file from each comparison was used to generate a table that uses gene names as the row names and the fold change as the values.

Gene set enrichment (GSE) is a method that generates all possible lists within a given range and assigns these lists to GO terms. It also considers fold-change values and calculates a p-value for each assigned GO term. In this study, gene set enrichment results were made for GO biological processes, GO molecular functions and GO cellular components for every comparison and the results were filtered according to their p-values; GO terms with a p-value higher than 0.05 were not taken into consideration.

The "clusterProfiler" (v4.4.4) R package was used for gene set enrichment analysis (Wu et al., 2021). This package also requires a database to be loaded, which searches for related GO terms, so org.Hs.eg.db database was also loaded into the R environment (*Bioconductor - Org.Hs.Eg.Db*, n.d.). gseGO() function was run with a minimum list length of 100, maximum list length of 500 as the list length parameters. The resulting tables were written into files.

2.4.1. Bringing GSE Results Together

The results tables were merged by considering the type of the project. The projects were separated into two main branches: Tissues, Cells. These were then further categorized by the tissue region and sex, or type of the cell, respectively.

Table 6 – Categorization of GSE results

Tissues	Cells
All	Astrocyte
Female	Microglia
Male	Neuron
Dorsolateral prefrontal cortex	Neural progenitor
Inferior parietal lobule	Cerebral organoid
Hippocampus	Induced Pluripotent Stem Cell
Middle temporal gyrus	
Olfactory Bulb	
Blood	
Lateral temporal lobe	

The GSE results were combined into a single file under these branches. As an example, all “female AD versus female healthy” GSE results were concatenated into a single table by adding rows from one table to another. Then, if there are rows containing the same gene, these rows are collapsed into one. Since there are also columns with numerical values, the average of them was taken while collapsing multiple rows into one.

For the 16 categories shown in the table (10 tissue, 6 cell culture categories), a combined GSE results file was generated for each 3 GO categories, leading to 48 files. All the categories were used for building a distance matrix to see the most distant samples by their GO profiles. Tissue categories were compared to cell culture categories to find similarities and differences among the projects. However, regional tissue categories were not included in this comparison as they would not produce meaningful results for the question we are trying to answer.

As explained above, the first three tissue categories were compared to the cell culture categories. The tables were merged (concatenated and collapsed) again during the comparison to find the intersection of the projects. As an example, while comparing the “female” table to the “astrocyte” table, the rows from the latter one were concatenated into the former one, and the genes that occurred multiple times were collapsed into one while the average was taken for the numerical values.

In addition to the intersection, the difference between the projects was also taken into consideration. Going over the same example, to find the “female - astrocyte” difference, genes that are already found in the astrocyte table were removed from the female table, and the opposite was made to find the “astrocyte - female” difference.

As a result, three tables were generated for each comparison: 1) A∩B 2) A-B 3) B-A – Leading to 54 comparison tables. These tables were used to identify the differences and similarities by GO terms and to build a distance matrix to numerically define the differences and similarities amongst the samples.

```
2
3  # Loading libraries
4     # These packages are installed via BiocManager::install()
5  library(clusterProfiler)
6  library(AnnotationDbi)
7  library(org.Hs.eg.db)
8
9  # Creating input
10    # The input consists of significant genes.
11    # Gene names or their Ensembl IDs will do.
12 data <- read.csv('PRJNA232669_DEresOrdered.csv',
13                   header = T, row.names = 1)
14
15 genes <- data$log2FoldChange
16 names(genes) <- rownames(data)
17 genes <- sort(genes, decreasing = T)
18
19 # GO Command
20 # All healthy vs all AD
21 gseGOBP <- gseGO(gene = genes,
22                  OrgDb = 'org.Hs.eg.db',
23                  keyType = 'ENSEMBL',
24                  minGSSize    = 100,
25                  maxGSSize    = 500,
26                  pvalueCutoff = 0.05,
27                  ont = 'BP') # BP: Biological processes
28 gseGOMF <- gseGO(gene = genes,
29                   OrgDb = 'org.Hs.eg.db',
30                   keyType = 'ENSEMBL',
31                   minGSSize    = 100,
32                   maxGSSize    = 500,
33                   pvalueCutoff = 0.05,
34                   ont = 'MF') # MF: Molecular function
35 gseGOCC <- gseGO(gene = genes,
36                  OrgDb = 'org.Hs.eg.db',
37                  keyType = 'ENSEMBL',
38                  minGSSize    = 100,
39                  maxGSSize    = 500,
40                  pvalueCutoff = 0.05,
41                  ont = 'CC') # CC: Cellular component
42
```

Figure 26 – Usage of gseGO() function from the clusterProfiler package. First, the results file from DESeq2 is loaded into the environment and is used to generate a table with gene names and fold change values of the genes. This table is then given into gseGO() for analysis.

2.4.2. GO Annotations and Analysis

The following analyses were made by outer and inner merging or subtracting the tables of groups shown in Table 6.

2.4.2.1 GO Term Topologies

A topological graph presenting the GO profiles of each condition is generated based on the tables containing the GO terms. Every table with BP, MF, and CC results was combined to get a complete list of GO terms. Then, every go term was assigned to a position on a grid with two axes. If a given condition exists on the table, such as a tissue table or a table of two samples intersecting, the position of that term is

marked with a ‘dot.’ Two features, the transparency and size, change depending on the number of genes representing that GO term. The square of that number is used for transparency, to make the difference in gene numbers more visible.

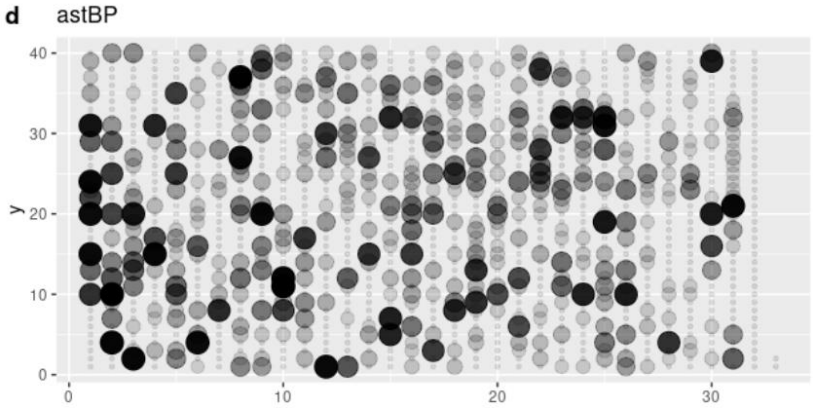

Figure 27 – BP topology graph of astrocyte, given as an example.

These plots allow for comparing different GO results visually as the term represented by each position is the same among all samples and comparisons.

2.4.2.2 Distance Matrix Analysis

In addition to the visual comparison, every sample was compared to each other using distance matrices built by the same principle as topology graphs to represent visual differences in numbers. It uses the number of genes in the gene set of the GO ID for every sample. If a GO ID is not found in the sample’s table, it is assigned a 0. From the values, Euclidean distance is calculated and the number is divided by 1000 to make the values easier to read.

Table 7 – Partial BP counts table.

ID	Description	allbp	fembp	malbp	astbp	mglbp
GO:0000018	regulation of DNA recombination	120	118.5	116	0	0
GO:0000045	autophagosome assembly	103.5	101	101	0	0
GO:0000070	mitotic sister chromatid segregation	164.33	157	159.5	163	164
GO:0000075	cell cycle checkpoint signaling	166	0	0	0	0
GO:0000077	DNA damage checkpoint signaling	0	0	0	0	0
GO:0000082	G1/S transition of mitotic cell cycle	184.33	0	182.5	0	0
GO:0000086	G2/M transition of mitotic cell cycle	132	0	125.5	0	0
GO:0000209	protein polyubiquitination	234	230.66	227.5	0	0
GO:0000280	nuclear division	413	395	364.5	0	361
GO:0000302	response to reactive oxygen species	176.4	0	179	0	173

2.4.2.3 Analysis with REVIGO

GO lists gathered from the GSE analysis were given to the online tool REVIGO (Supek et al., 2011) for visualizing the terms by their categories.

CHAPTER 3

RESULTS

3.1. Differentially Expressed Genes Analysis

A total of 19 projects were analyzed with DESeq2 for DEG analysis, leading to 42 comparisons. A gene was considered significant if its adjusted p-value was below 0.05. The number of significant genes for each comparison is shown in Table 8.

Table 8 – Number of significant genes by types, subtypes, and comparisons.

Project ID	Type	Subtype	Comparison	Significant Genes (p.adj<0.05)
232669	Tissue	Dorsolateral prefrontal cortex	AD vs C	0
279526	Tissue	hippocampus	AD vs C	71
			femAD vs femC	10
			mAD vs mC	147
399530	Cell	iPSC derived brain cell	APOE3 vs APOE4	10
		iPSC derived astrocyte	astAPOE3 vs astAPOE4	631
		iPSC	ipsAPOE3 vs ipsAPOE4	298
		iPSC derived microglia	mglAPOE3 vs mglAPOE4	847
		iPSC derived neuron	neuAPOE4 vs neuAPOE4	339
451437	Tissue	Olfactory bulb	AD vs C	1
482601	Cell	iPSC derived cerebral organoid	APOE3 vs APOE4	4521
527202	Cell	iPSC derived neuron	APP vs WT	9646
			PSEN1 vs WT	8237
			APP-PSEN1 vs WT	6488
559812	Cell	iPSC derived microglia	PSEN1 vs PSEN1corrected	2
576835	Cell	iPSC derived astrocyte	PSEN1 vs PSEN1corrected	2308
			femPSEN1 vs femPSEN1corrected	2833
			mPSEN1 vs mPSEN1corrected	5657
603192	Tissue	Inferior parietal lobule	APOE2 vs APOE3	8514
			femAPOE2 vs femAPOE3	5789
			mAPOE2 vs mAPOE3	1298

			APOE2 vs APOE4	9906
			femAPOE2 vs femAPOE4	3677
			mAPOE2 vs mAPOE4	5291
			APOE3 vs APOE4	3
			femAPOE3 vs femAPOE4	6157
			mAPOE3 vs mAPOE4	1
643561	Cell	hESC derived microglia	SORL1-KO vs WT	819
			SORL1-VR vs WT	448
			TREM2-KO vs WT	812
			TREM2-VR vs WT	2555
644383	Tissue	Lateral temporal lobe	AD vs C	3660
662330	Cell	iPSC derived microglia	TREM2-KO vs WT	4254
675864	Tissue	blood	AD vs C	22
683625	Cell	Neural progenitor	AD vs C	3537
688060	Tissue	Middle temporal gyrus	AD vs C	44
688885	Cell	iPSC derived cerebral organoid	serumTrt vs C	189
714081	Tissue	blood	AD vs C	3
727602	Tissue	hippocampus	AD vs C	5031
			femAD vs femC	4098
			mAD vs mC	35
767074	Tissue	hippocampus	AD vs C	18

The scale for significant genes varies between 0 to 9906. Different projects with similar designs showed a wide variety, such as PRJNA727602 and PRJNA76704. Both of these studies are on the hippocampus region, where one study provides 5031 significant genes, and the other has 18 significant genes without considering the 'sex' variable (Table 8).

Blood samples (PRJNA675864, PRJNA714081) and olfactory bulb (PRJNA451437) studies were included as references to test our methodology, meaning that the number of significant genes expected from them is close to zero. The studies have 22, 3, and 1 significant gene(s). These observations supported our methodology. The percentages of significant genes are calculated by adding the number of significant genes to show how many genes each project represents (Figure 28).

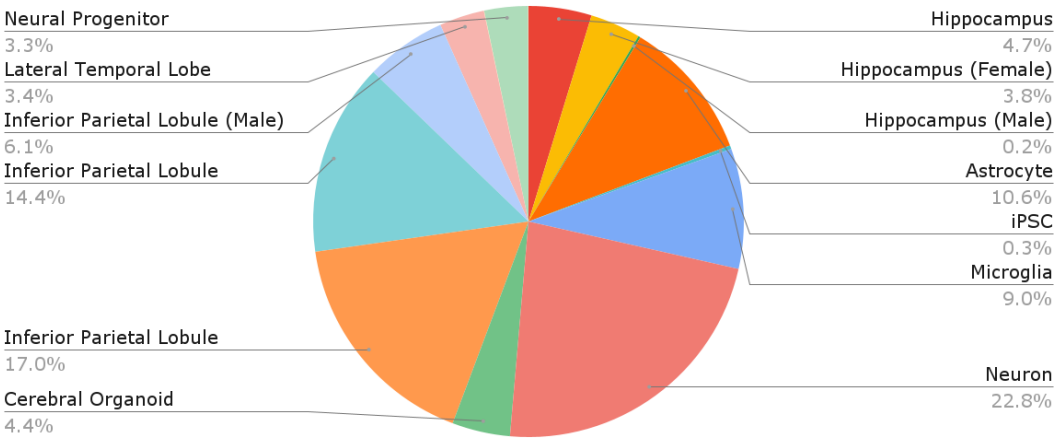

Figure 28 – Percentage of significant genes by sample group. These are aggregated values, adding the number of significant genes to show how many genes each project represents.

Analysis of the rest of the post-mortem tissue studies, PRJNA232669 (dorsolateral prefrontal cortex), PRJNA279526 (hippocampus), PRJNA688060 (middle temporal gyrus), and PRJNA767074 (hippocampus) showed relatively low differentiation, which is unexpected since these are the regions where AD neuropathology are reported. This may be due to experimental errors or problems in tissue preservation. In any case, while considering the data for comparison and taking the union of samples with common features, these data were included, not discarded.

Tissue samples with a high number of significant genes are PRJNA603192 (inferior parietal lobule), PRJNA644383 (lateral temporal lobe), and PRJNA727602 (hippocampus). The inferior parietal lobule study models all types of APOE and most of the significant genes are observed between ε2 vs. ε3 variants. A high number of significant DEGs are revealed in female samples, suggesting different pathways to be considered in female individuals with AD. A similar result is observed in PRJNA727602 (hippocampus): 4098 significant genes in females, compared to 35 genes in males (Figure 29).

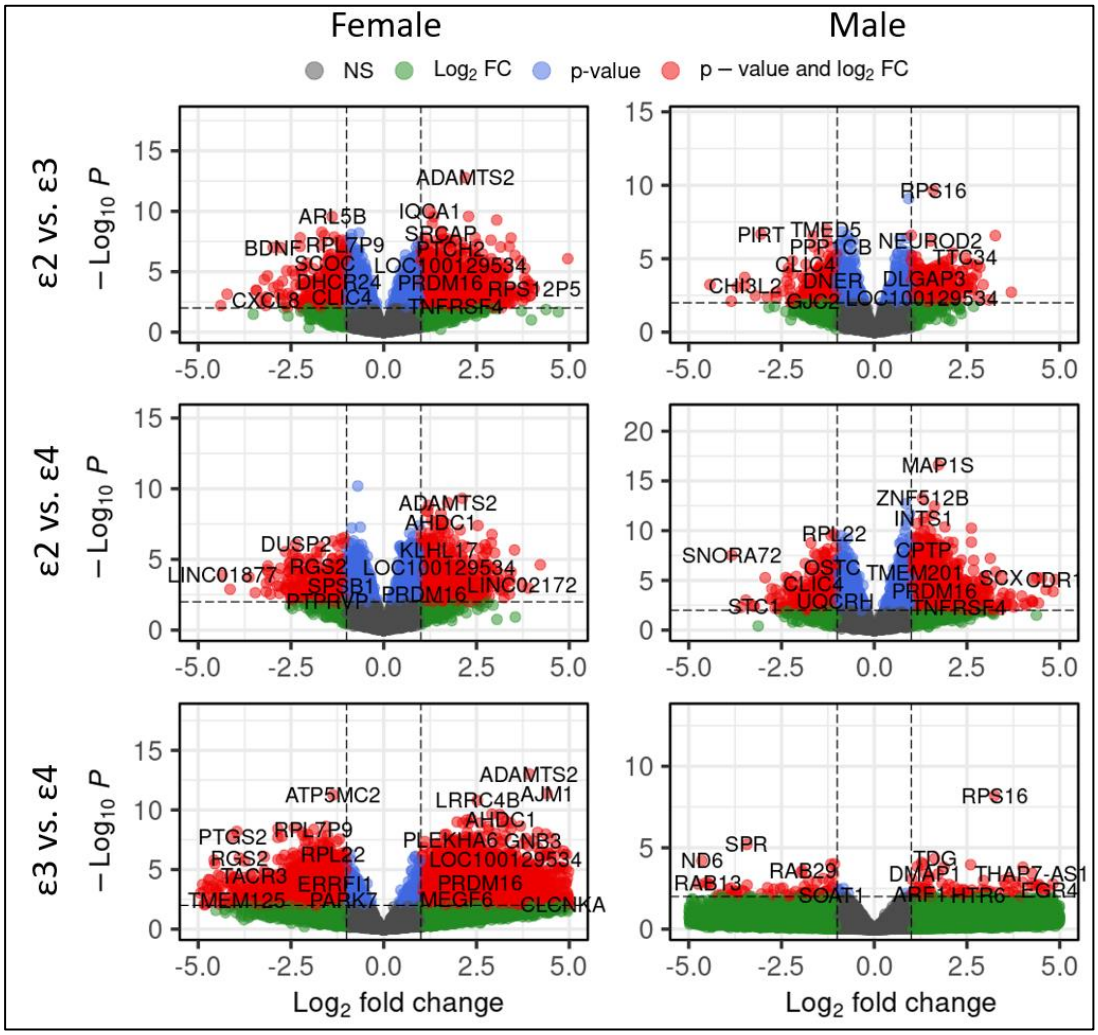

Figure 29 – Volcano plots of PRJNA603192 samples given as an example. Female samples provide richer results in terms of significant genes. ε3 vs. ε4 results for male has RPS16, which is the only gene that surpasses the adjusted p-value threshold.

Cell studies, on the other hand, provide a relatively high number of significant genes in contrast to tissue studies, the lowest number being around 300. The only study showing low differentiation was PRJNA688885 (iPSC-derived brain organoids), which shows 189 significant genes. This study creates an extracellular environment instead of editing AD-related genes to risky variants, which may be the reason for low differentiation.

3.2. Gene Set Enrichment Analysis

As there are many sets with a low number of significant genes, directly converting them to GO terms would mean further decreasing the number of projects. To include them in the downstream analysis steps, GSE was selected as the best approach as it would include fold-change values, too. Further information for each dataset can be found in Appendix B at the DEG level and Appendix C at the GSE level.

GSE was run for every comparison shown in Table 8. We observed that inferior parietal lobule (IPL) ε3 vs. ε4 results for males had the most number of enriched GO terms when compared to ε2 vs. ε4 male samples, even though the former one has more significant genes.

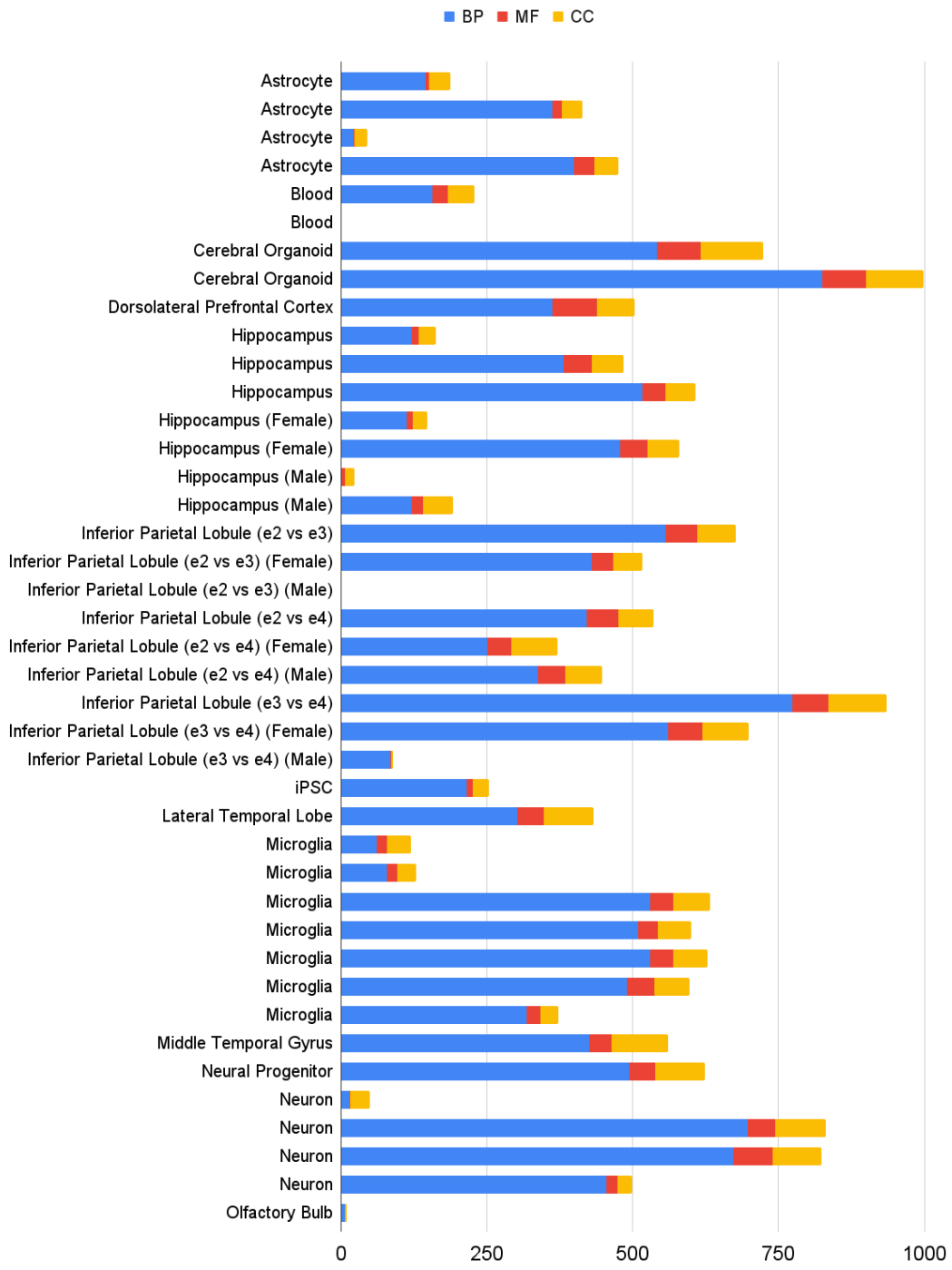

Figure 30 – Number of enriched GO terms by samples in each project.

3.3. Distance Matrix Analysis

Just like grouping tissues under sex type, we can do the same for brain regions to build a distance matrix. According to the distance numbers, the reference cell type iPSC is showing a profile very similar to the reference tissue types blood and olfactory bulb (Figure 31).

The highest amount of distance can be observed between neurons and the olfactory bulb, while the most similar samples are neural progenitor cells and the middle temporal gyrus. This region, along with the hippocampus and interior parietal lobule, shows the highest amount of similarity to cell samples. All cell types show similarity to these regions and distance to other tissues, except for astrocytes and iPSC. While iPSCs are showing the reverse profile in comparison to those cells, astrocytes are always in the middle-ground in terms of distance (Figure 31).

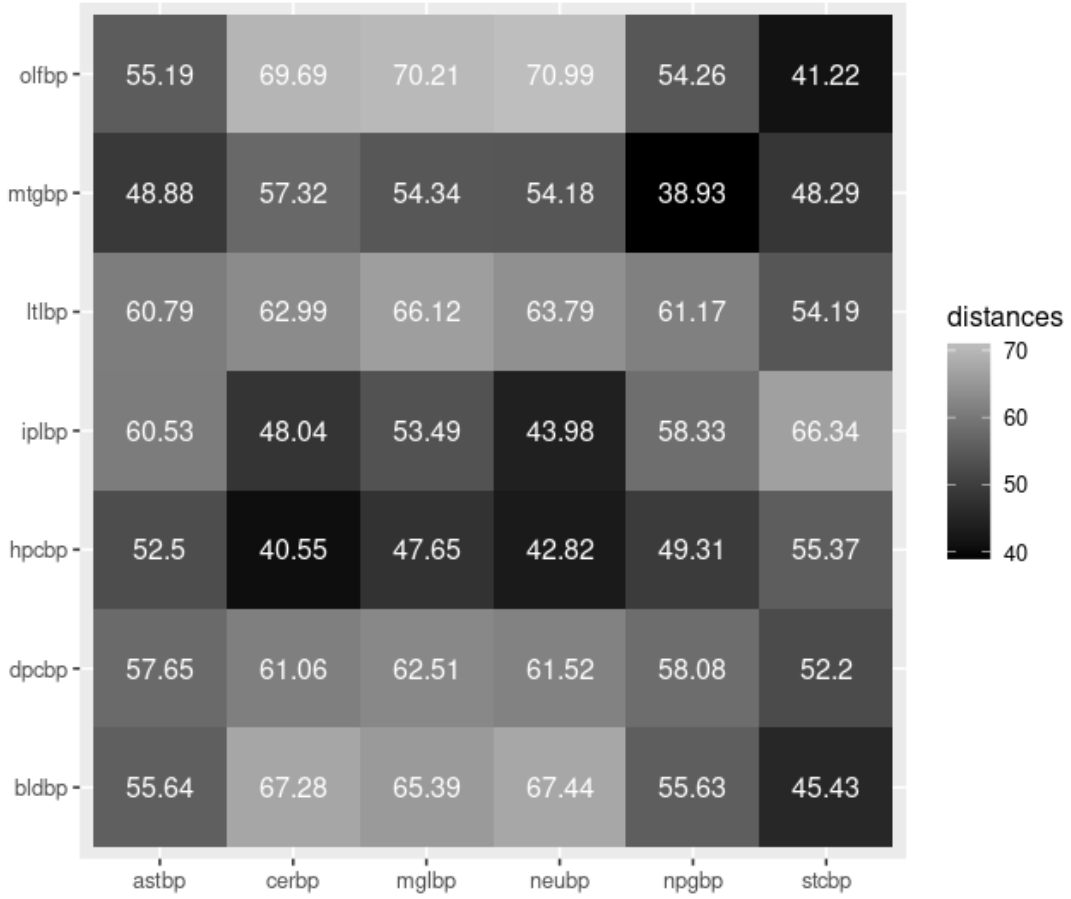

Figure 31 – Distance matrix of cells vs. brain regions by BP. olf: Olfactory bulb, mtg: Middle temporal gyrus, ltl: Lateral temporal lobe, ipl: Inferior parietal lobule, hpc: Hippocampus, dpc: Dorsolateral prefrontal cortex, bld: Blood, ast: Astrocyte, cer: Cerebral organoid, mgl: Microglia, neu: Neuron, npg: Neural progenitor, stc: iPSC

iPSC samples are considered as the reference, providing a background signal. Neural progenitor cells are not the main subjects of AD studies. Cerebral organoid samples are used to see if they have the most similarity with tissue samples, as they are a mixture of differentiated brain cells, as in tissues (Figure 32). Considering these,

astrocytes, neurons, and microglia were selected for comparison with tissue samples, as they also are the main subject of AD research.

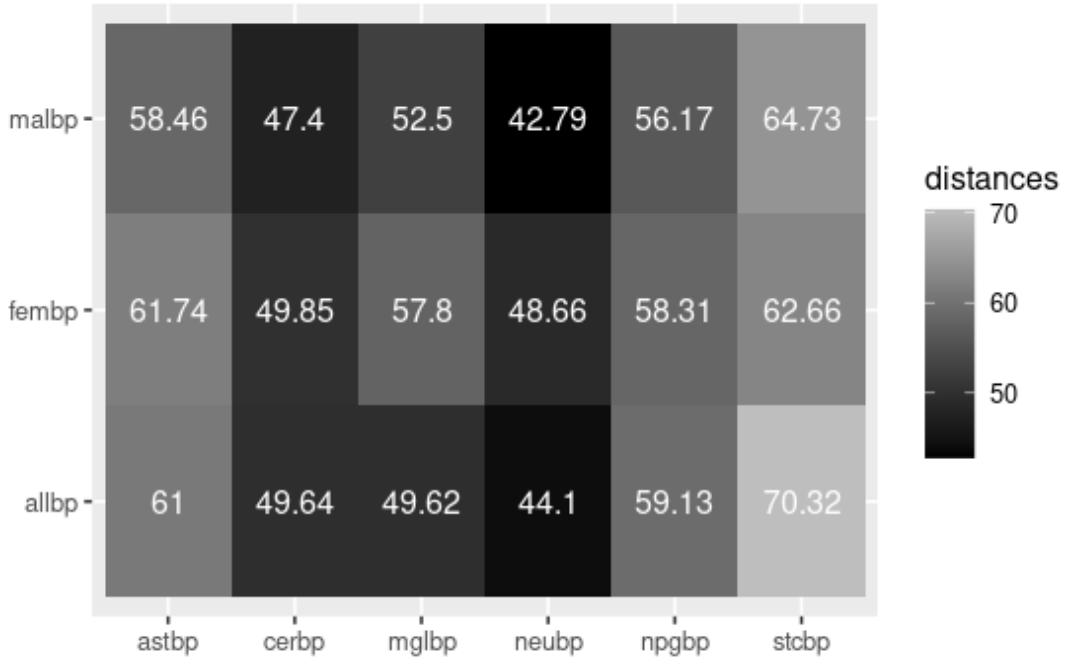

Figure 32 – Distance matrix of cell samples vs. tissues by the number of genes in each BP term. mal: Male, fem: Female, all: Tissue samples excluding sex variable, ast: Astrocyte, cer: Cerebral Organoid, mgl: Microglia, neu: Neuron, npg: Neural progenitor, stc: iPSC

3.4. GO Topology Analysis

We suggested a new approach to visualize the GO terms revealed with the topology graphs, which summarize the GO annotations and emphasize the differences between sample sets. The list of GO terms with their gene set size was used to build the topology graph for each category as described in the methods sections. Biological process (BP) is represented in Figure 33, Molecular Function (MF) in Figure 34, and Cellular Component (CC) in Figure 35. The graphs do not represent all GO terms but every GO term assigned to our data. Every position represents a different GO term, which is at the same position among all graphs.

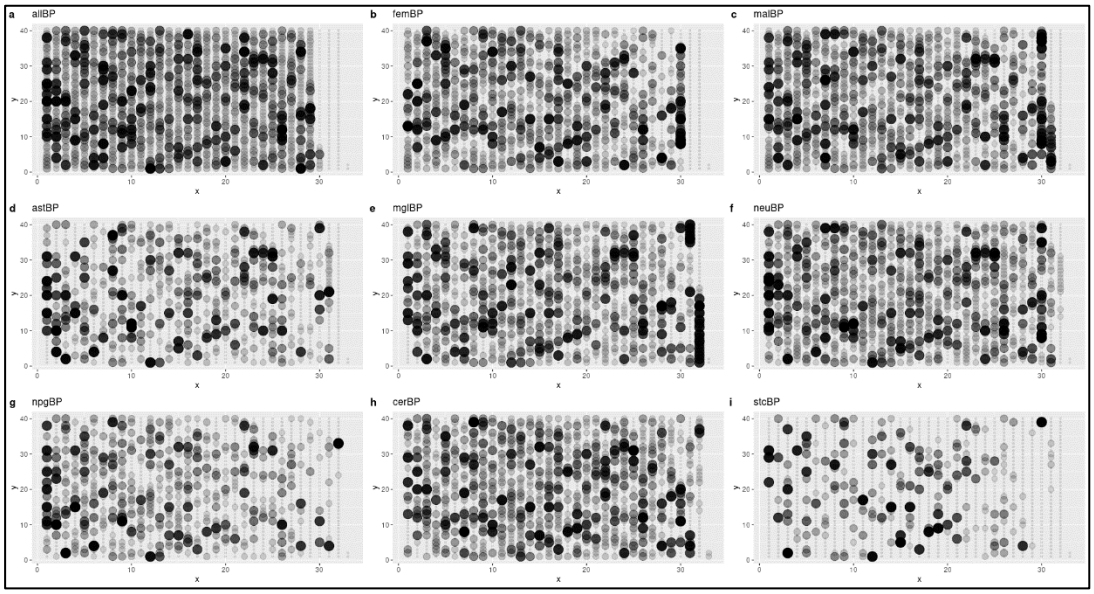

Figure 33 – BP topologies of major sample groups. a: Tissue samples without sex variable. b: Tissue samples from females. c: Tissue samples from males. d: Astrocyte samples. e: Microglia samples. f: Neuron samples. g: Neural progenitor cell samples. h: Cerebral Organoid samples. i: iPSC samples.

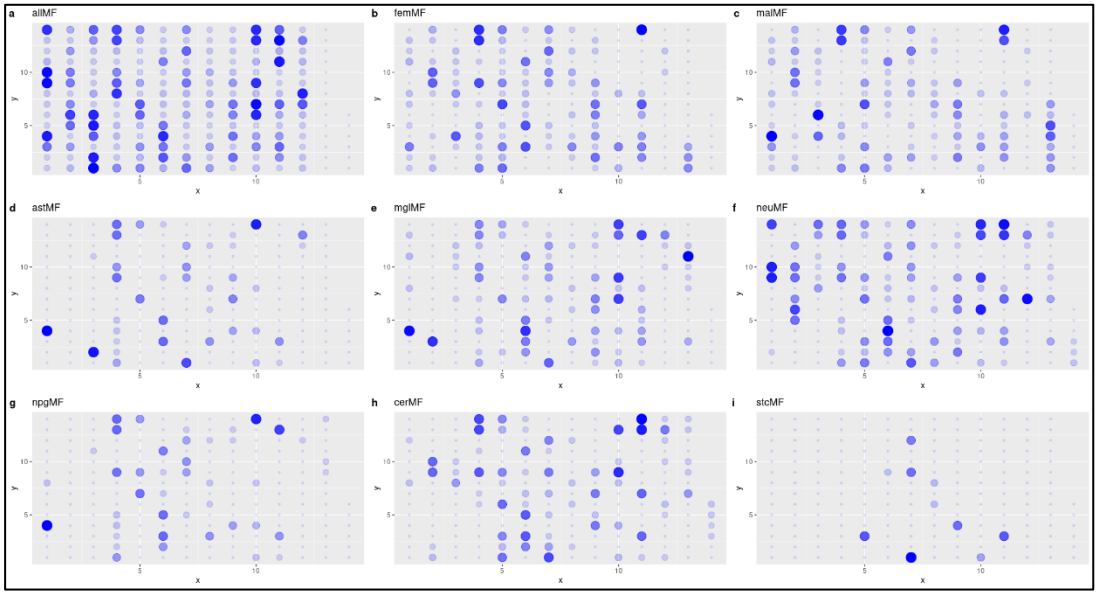

Figure 34 – MF topologies of major sample groups. a: Tissue samples without sex variable. b: Tissue samples from females. c: Tissue samples from males. d: Astrocyte samples. e: Microglia samples. f: Neuron samples. g: Neural progenitor cell samples. h: Cerebral Organoid samples. i: iPSC samples.

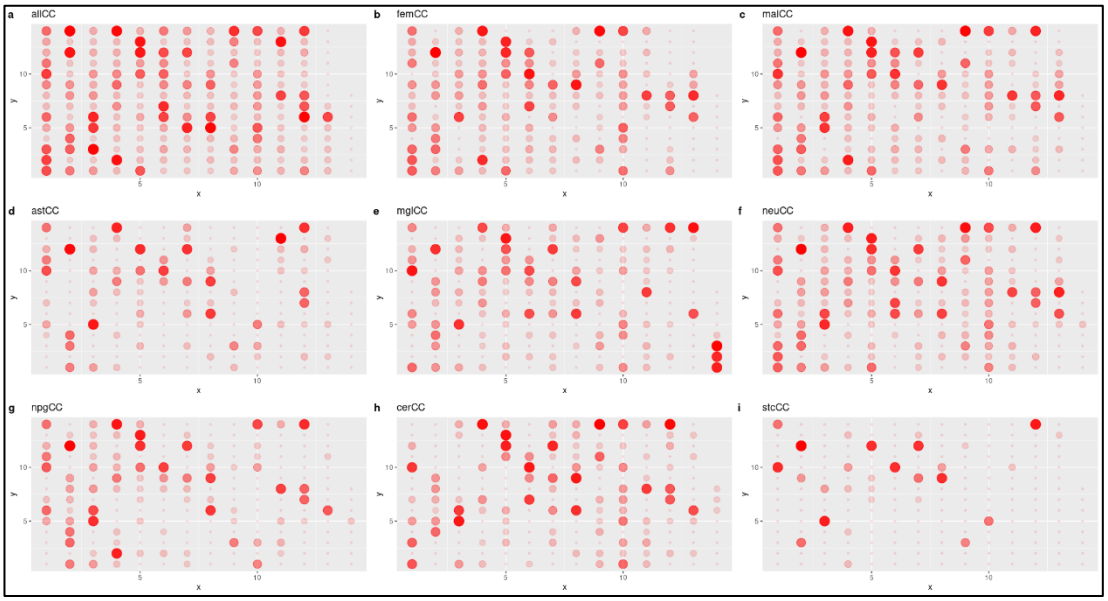

Figure 35 – CC topologies of major sample groups. a: Tissue samples without sex variable. b: Tissue samples from females. c: Tissue samples from males. d: Astrocyte samples. e: Microglia samples. f: Neuron samples. g: Neural progenitor cell samples. h: Cerebral Organoid samples. i: iPSC samples.

Our analysis showed no specific GO terms are observed in tissue samples but in microglia (Figure 33). Neurons and cerebral organoids presented similar profiles, yet the signals were not as strong as in microglia. Like 'blood' samples in the tissue group, iPSC samples were used as a reference in the cell group. As expected, they provide the scarcely packed topology, followed by astrocytes and neural progenitor cells (Figure 33).

In summary, neurons, microglia, and cerebral organoid tend to show similar results to tissue samples. Astrocyte, neural progenitor, and iPSC samples are similar, while they are the most distinct in comparison with tissue samples.

Some common patterns can be seen among all samples, with tissue samples without sex variable with the highest signal. Female and male samples are very similar, especially for MF and CC graphs (Figures 34 and 35).

MF and CC graphs were also informative, but in contrast to BP graphs, they did not provide any additional information, creating redundancy. Due to this, most other comparisons are made using BP topology graphs. BP terms were also used to build the distance matrices shown in this section and Appendix C.

3.4.1. Astrocytes

In comparison to tissues, astrocytes show the most distinct transcriptomic profile. The intersection graphs show similar groupings of BP IDs. The IDs present in all three astrocyte/tissue difference graphs show non-similar topologies, especially on the leftmost part. This also shows male and female samples show differences in BPs that are affected in astrocytes when these is an AD pathology (Figure 36).

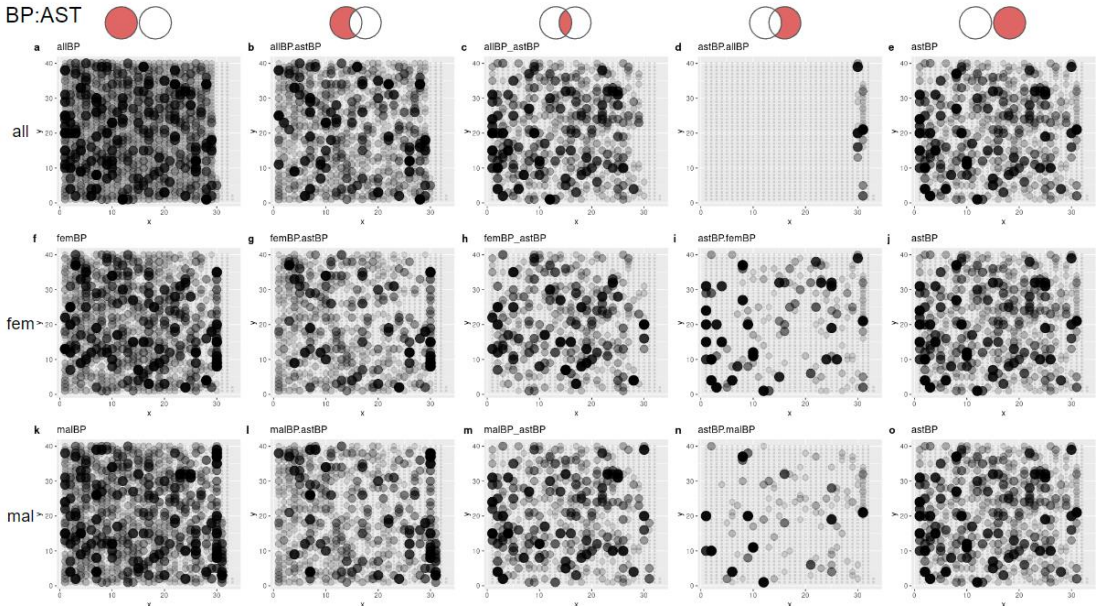

Figure 36 – Astrocyte topologies in comparison to tissues. The first column shows GO term topologies for tissue samples. The second column shows BP terms enriched in tissues but not in astrocytes. The third column is the common enriched BP terms in astrocytes and tissues. The fourth column shows BP terms enriched in astrocytes but not in tissues. The last column shows astrocyte topologies. All: Tissue samples excluding sex variable, fem: Female tissues, mal: Male tissues

3.4.2. Microglia

Microglia share a high amount of similarities to tissues in terms of BP IDs. This can be observed in the intersection of the topology graphs (Figure 37). Tissue sample – cell sample difference is much lower in contrast to astrocytes. On the other hand, cell sample – tissue sample differences show more similarities between female and male samples, yet, there are still more biological processes that are missing from the female samples.

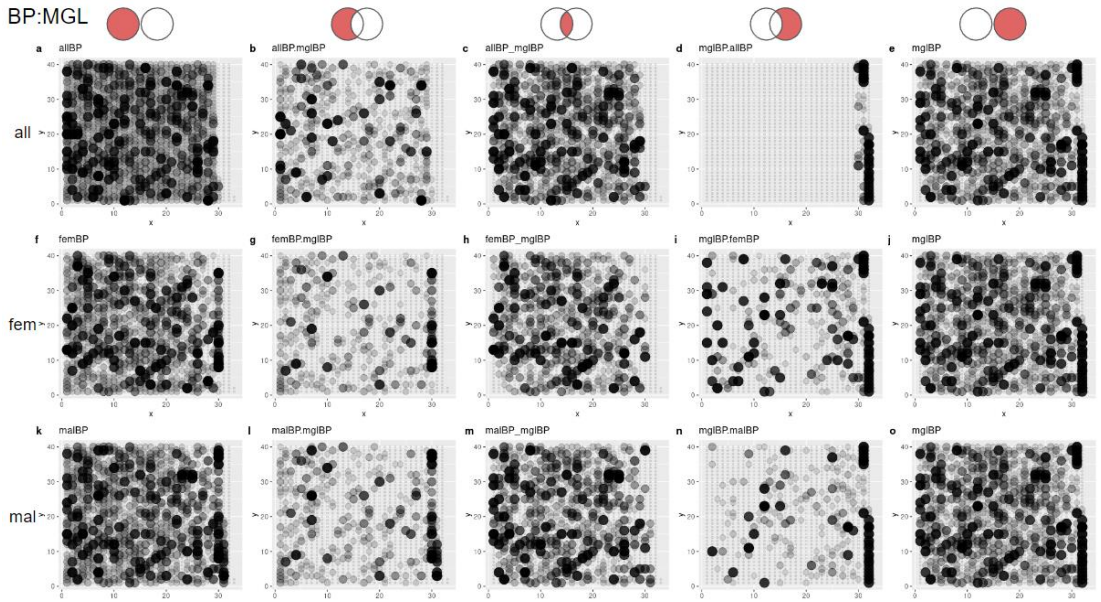

Figure 37 – Microglia topologies in comparison with tissues. The first column shows tissue sample topologies. The second column shows BP terms that are enriched in tissues but not in microglia. The third column is the common enriched BP terms in microglia and tissues. The fourth column shows BP terms enriched in microglia but not in tissues. The last column shows microglia topologies. all: Tissue samples excluding sex variable, fem: Female tissues, mal: Male tissues

3.4.3. Neurons

The results of neuron – tissue comparisons show a high amount of similarity to microglia results in differences and intersections (Figure 38). Female tissue samples tend to have more missing BPs again, the intersection of topologies is very crowded, and the results show similarity between female and male samples.

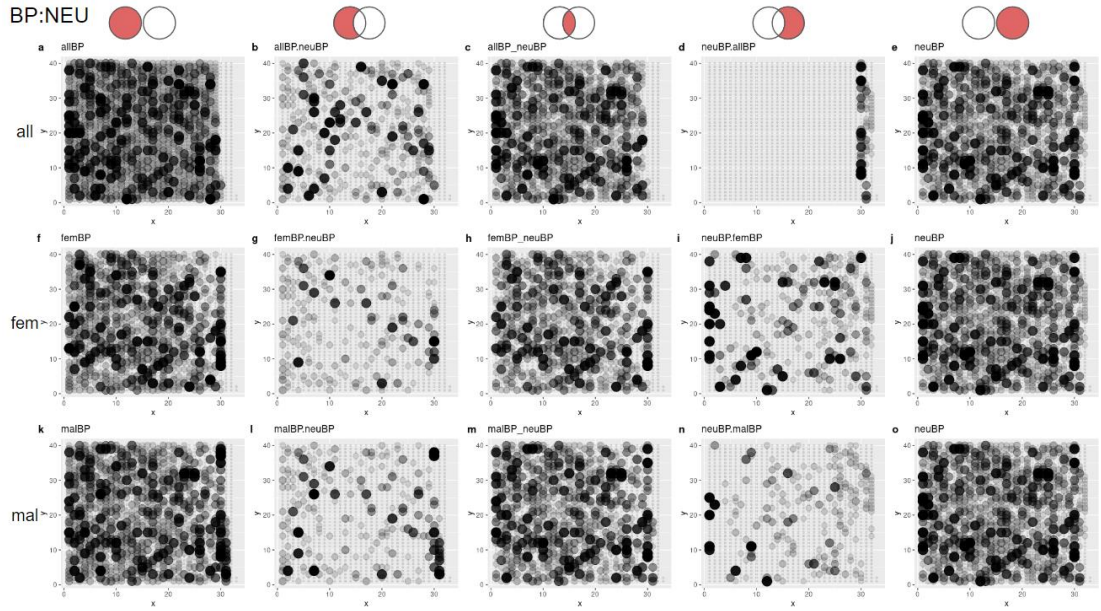

Figure 38 – Neuron topologies in comparison with tissues. The first column shows tissue sample topologies. The second column shows BP terms that are enriched in tissues but not in neurons. The third column is the common enriched BP terms in neurons and tissues. The fourth column shows BP

terms enriched in neurons but not in tissues. The last column shows neuron topologies. all: Tissue samples excluding sex variable, fem: Female tissues, mal: Male tissues

According to topology graphs and distance matrices analysis, iPSC profiles are the most distant to tissue samples. Neurons show the highest similarity to tissues, followed by cerebral organoid (Figure 31). As seen in Figure 31, astrocytes show similarity to iPSCs, distance to tissue samples, and numbers only slightly higher than neural progenitors. For microglia, we can see that it is most similar to neurons but less distant to tissues than other cell types (Figure 31). These tissue-cell comparison results are also reflected in the cell-cell distance matrix (Figure 39).

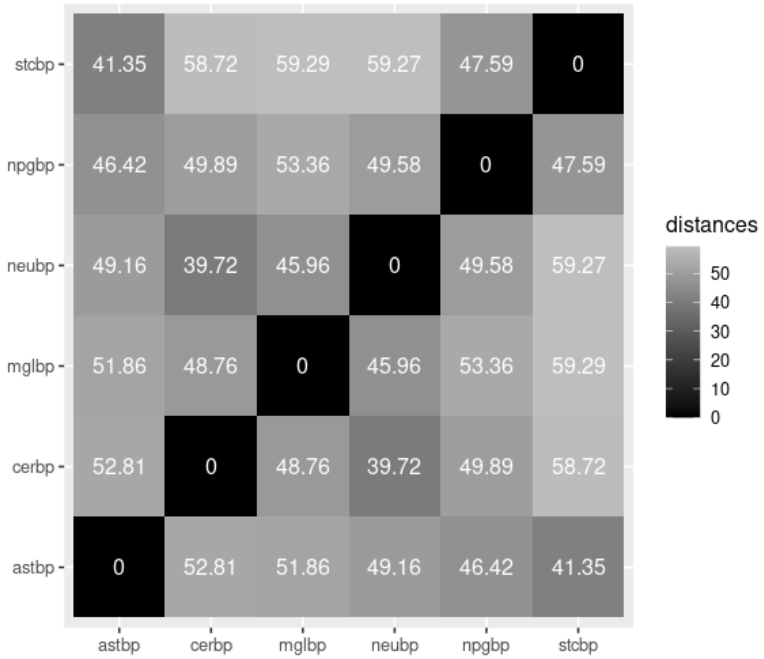

Figure 39 - Distance matrix of cell models by BP. ast: Astrocyte, cer: Cerebral organoid, mgl: Microglia, neu: Neuron, npg: Neural progenitor, stc: iPSC

3.4.4. Comparisons of Tissue Groups

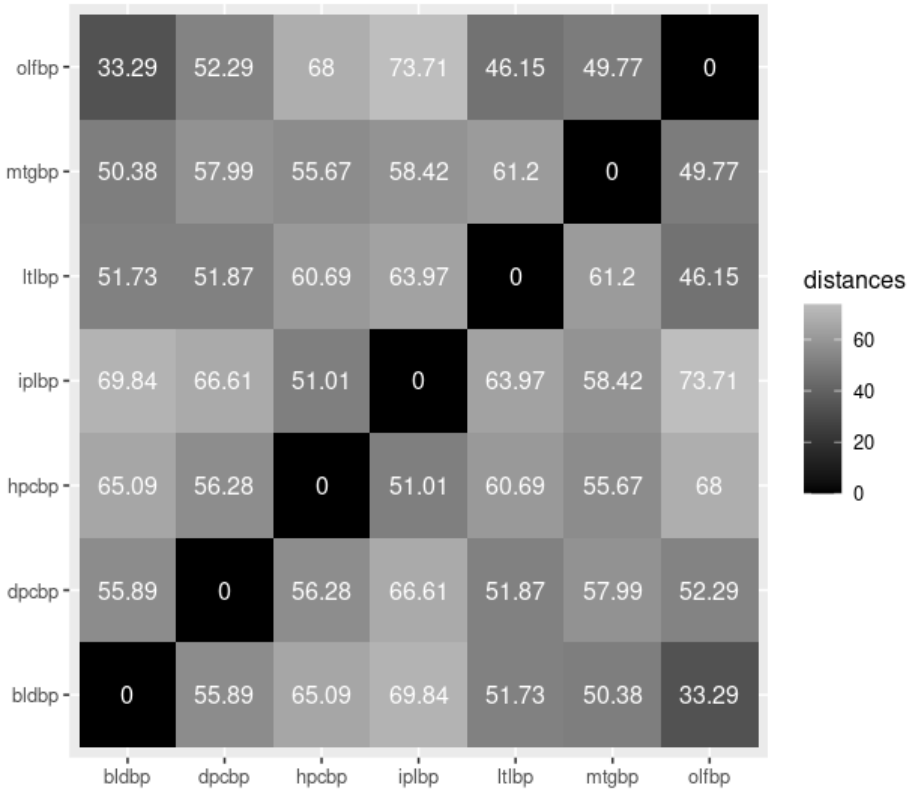

Figure 40 - Distance matrix of brain regions by BP. olf: Olfactory bulb, mtg: Middle temporal gyrus, ltl: Lateral temporal lobe, ipl: Inferior parietal lobule, hpc: Hippocampus, dpc: Dorsolateral prefrontal cortex, bld: Blood

When distances are considered, the reference regions showed the highest similarity: blood tissue and olfactory bulb samples. If we take the minimum distance and 73,71 as the maximum (interior parietal lobule vs. olfactory bulb), our scale is between 33,29 to 73,71. Considering this, most of the scores are on the upper part of this scale, but the scores are very similar, meaning each brain region is about the same distance from the other. The only exception was the LTL and olfactory bulb comparison, which is expected due to low scores of LTL results. Tissue sample comparisons may contain a bias due to some data having less reliable results (DPC, LTL) (Figure 40). Overall each sample has its own set of unique GO terms, while the number of enriched GO terms is close in every sample.

When we compare the regions with sample groups separated by sex (Figure 41), we see female dominance in terms of similarity compared to cell samples. Blood tissue and especially olfactory bulb sample results are very distant, followed by LTL and DPC samples. The highest similarity is observed in the inferior parietal lobule (IPL), hippocampus (HPC), and middle temporal gyrus (MTG), respectively. This order only represents that the tissue samples are mostly pooled from these. The IPL-male similarity is also very high, probably due to a sampling bias: LTL and DPC do not provide beneficial results, MTG design does not include sex information, and there

are a small amount of HPC results for male samples, leaving the male pool with IPL samples mostly.

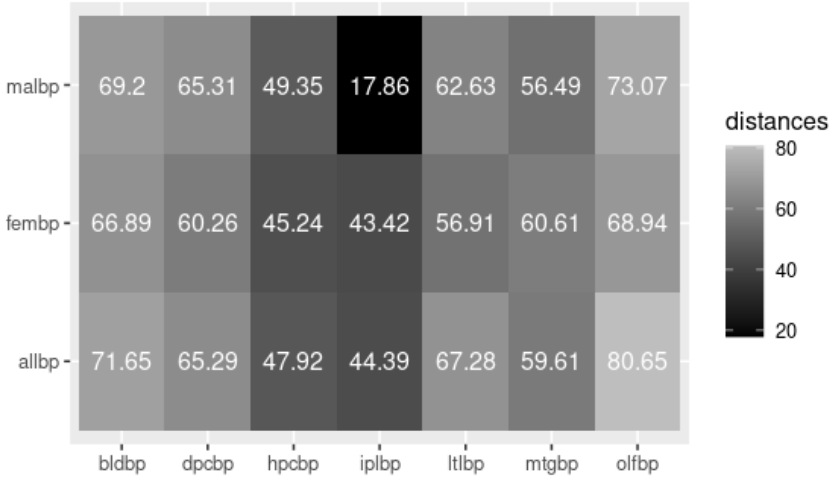

Figure 41 – Distance matrix of brain regions vs. sexes by BP. mal: Male samples, fem: Female samples, all: Tissue samples excluding sex variable, bld: Blood, dpc: Dorsolateral prefrontal cortex, hpc: Hippocampus, ipl: Inferior parietal lobule, ltl: Lateral temporal lobe, mtg: Middle temporal gyrus, olf: Olfactory bulb

3.4.5. Comparison of Cell Types

The maximum distance and average scores of the cell-to-cell distances were lower compared to tissue-to-tissue distances. The higher similarity probably arises from the experimental design. All of these cells are provided with similar factors for conversion from iPSC, and they have a strictly controlled environment compared to the real human brain.

The topology plots have shown us the similarities between neurons, microglia, and cerebral organoid, in addition to the similarities between astrocytes, neural progenitors, and iPSCs. We observed that our reference, iPSCs, are showing the highest amount of distance while being most similar to astrocytes and then neural progenitor cells. The highest similarity was observed between cerebral organoids and neurons, which suggests the cell content of cerebral organoids mostly consists of neurons, or the extracellular environment and cellular communication creates a transcriptomic profile in other cells that resembles neurons. Similarly, the topology plots have shown us the similarities between neurons, microglia, and cerebral organoid, in addition to the similarities between astrocytes, neural progenitors, and iPSCs.

As seen in Figure 32 and Figure 39, astrocytes do not show a strict pattern in contrast to brain regions or even tissue samples by sex and keep a similar distance to all of them. This suggests that the biological processes affected by AD in astrocytes can lead to a pathology that can be observed in every region of the cortex and is similar in both females and males. The low similarities suggest that astrocyte pathology is not a strong driver as microglia or neuron pathology.

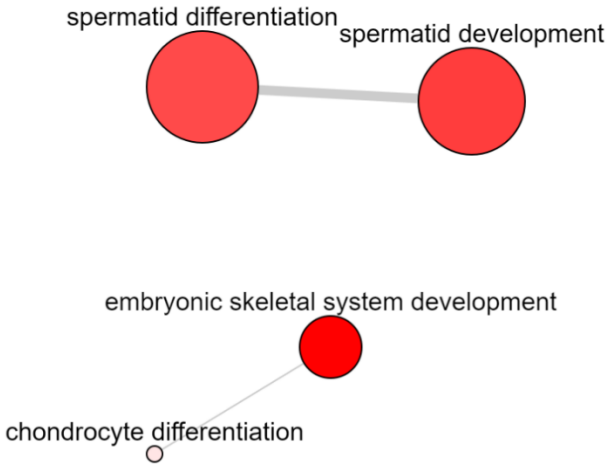

Figure 42 - Biological processes that are affected in AD, specific to astrocytes. The size of the bubbles represents the number of genes related to that GO term in the Homo sapiens GO database. And the color represents the p-value: The lighter the color, the more significant the hit.

The biological processes specific to astrocytes might be the artifacts formed when cell differentiation and proliferation were taking process (Figure 42), but they also might be related to pathways that are used by astrocytes to provide energy to the synaptic activity of neurons, which are also seen in developing germ cells (Matos e al., 2021). Overall, according to the distance matrices and the specific BPs, in AD studies, astrocytes do not represent the tissues as well as other cells. In brain tissues, astrocytes are connected to thousands of neurons while also delivering nutrients and providing information to surveying microglia through cell signaling. Our suggestion would be to use astrocytes in conjunction with other cell types if their transcriptomic profile is the topic of research and cell models are to be used.

Microglia show more distant results than neurons. As they are affected externally during an AD pathology, instead of internally like neurons, this result is expected (Figure 32).

When we focus on the differences between the sexes, again, we see that these cells are more distant to females than males (Figure 32).

According to recent studies, a different version of C3 complementary protein, which activates microglia, is more abundant in females (Yang et al., 2022). And the activation of microglia occurs when they are exposed to external elements. In this case, the activation levels in female and male brains and the cell culture should be considered. Microglial modeling with gene editing techniques may not represent the level of activation observed in human, especially female, brains (Figure 43).

Figure 43 - Biological processes that are affected in AD, specific to microglia. The size of the bubbles represents the number of genes related to that GO term in the Homo sapiens GO database. And the color represents the p-value: The lighter the color, the more significant the hit.

In terms of distinct BPs, microglia show similarities to tissue samples as these BPs are related to cell signaling, development, and transport. This may be due to the surveying nature of microglia. A microglial cell is always in communication with the extracellular environment. For this to happen, a lot of transportation and signaling have to occur. As AD disrupts normal functions of primary brain cells, it is expected that these are pathways to be disrupted in microglia (Figure 43).

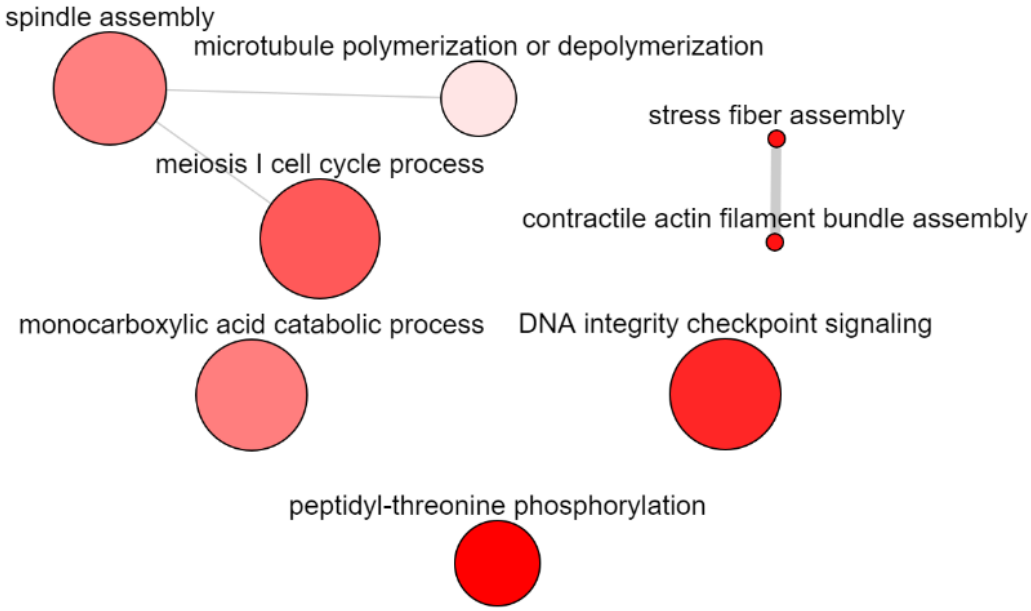

Figure 44 - Biological processes that are affected in AD, specific to neurons. The size of the bubbles represents the number of genes related to that GO term in the Homo sapiens GO database. And the color represents the p-value: The lighter the color, the more significant the hit.

The BPs specific to neurons are mostly structural, such as fiber assembly and microtubule polymerization (Figure 44). The underlying reason can be from the cell differentiation with a high chance. Neurons are not as flexible as microglia or astrocyte; they have to form stable connections, which is provided by stable, fibrous extensions. These processes might be disrupted in AD samples. In brain tissue, when AD develops to a pathological state, existing neurons are matured, and the only flexibility is introduced to them by surrounding astrocytes and microglia. Yet, in a petri dish, the neurons are young and forced to develop. Our suggestion is the experimental methodology creates these differences.

3.4.5.1 Common Pathways in Neurons and Microglia

In the common pathways of neurons and microglia, which are distinct from tissue samples and astrocytes, we can see filament-based movement and ERBB signaling pathway with the highest significance (Figure 45).

Filament-based movement may be related to the neurons due to the migration of newly generated neurons found in brains. Also, filament structures might be related due to reasons explained under the previous title. The relation to the microglia is obvious as they are very active cells with a high amount of movement. This may also be related to neuron-microglia communication that induces the activation and movement of microglia.

ERBB pathways play a role in astrogenesis during development (Buonanno & Fischbach, 2001). Previous studies in our group have shown the enrichment of

genes with LOAD-associated variants, specifically in ERBB4 pathways (Sevda Rafatov, METU thesis 2022). ERBB4 signaling through the nucleus promotes various transcription factors and acts as a transcription co-factor. In neuronal cells, ERBB4 is a co-factor of inhibition of expression of astrocyte differentiation genes GFAP and S100B (Sardi et al., 2006). ERBB4 s80 was also shown to bind activated estrogen receptors in the nucleus and act as its transcriptional co-factor in promoting transcription of some estrogen-regulated genes, such as progesterone receptor gene NR3C3 and CXCL12, i.e., SDF1 (Zhu et al., 2010). ERBB4 s80 is also able to translocate to the mitochondrial matrix and act as a pro-apoptotic factor (Naresh et al., 2006).

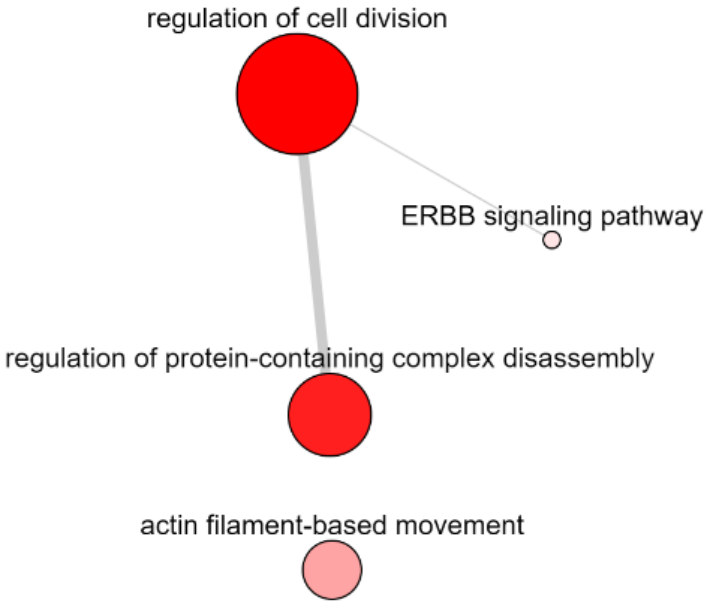

Figure 45 – The common biological processes that are affected in AD, in both microglia and neurons. The size of the bubbles represents the amount of genes related to that GO term in Homo sapiens GO database. And the color represents p-value: The lighter the color, the more significant is the hit.

CHAPTER 4

DISCUSSION

Most of the causes of AD development are still not understood, and research shows many environmental and genetic factors behind the scenes. It should be noted that the models are generated based on our current understanding of AD, and various cell types are tested for different variants in AD-related genes. Tissue sample designs do not always contain sex variables; they can come from different brain regions and may not provide AD-related results. The results presented here can improve or change with the addition of new datasets. The following discussion is presented considering these limitations of the experimental design.

In tissue studies with sex variables, female samples show more DEGs than males. This observation supports the effect of the sex-related pathways, as women have a higher risk of developing AD (Scheyer et al., 2018).

In contrast to tissue samples, almost all cell studies provided a high number of DEGs. The difference probably arises from the controlled environment available in cell studies and tissue samples going through preservation steps that may interfere with the transcriptomic profile of the cells. The samples might have different pathological distributions, or on different stages, so they are not precisely matched. All these variables are eliminated in cell models.

The distance matrices showed that the middle temporal gyrus (MTG), inferior parietal lobule (IPL), and hippocampus (HPC) had the highest amount of similarity to the cell profiles. The dorsolateral prefrontal cortex (DPC) and lateral temporal lobe (LTL) should also be providing similar results as they are target regions in AD development, yet we see that they are more distant, having values above 60. This may be related to the low number of significant genes in these regions, shown previously in Table 8. The gene sets may have led to different GO term annotations, making these regions distant. MTG, IPL, and HPC were similar to the cell models that ended up with intersecting gene sets and GO term annotations. However, the low number of significant genes in DPC and LTL does not affect the “total tissue samples vs. cell types” comparison since all tissue samples were used to generate an outer join table instead of an inner join table.

When enriched GO terms of tissues from male and female samples are compared, the samples from male tissues resemble the cell models. This was the case for neural progenitors, astrocytes, and cerebral organoids. The distance was highest for

neurons, closely followed by microglia, the primary cell types responsible for AD development.

Considering the number of DEGs in female and cell samples, we'd expect to see a higher similarity than male-cell comparison as higher numbers in the same disease would lead to higher intersections. This possibly indicates that the significant DEGs in females arise from different pathways that cannot be directly explained by brain tissue metabolomics, and the effect of the female physiology is not reflected in the cell samples. This suggests that cell model studies cannot equally represent the effect of sex on AD pathology.

Distinct metabolic pathways are observed in female and male tissue samples in contrast to cell models. In female samples, ion transport pathways, nucleic acid recombination, response to the environment, and T-cell activation regulation are differentially regulated in AD (Figure 46). Male samples showed differentiation in DNA repair, autophagy, mitotic cell cycles, cell phase transitions, molecule transportations, and nucleic acid catabolic processes (Figure 47). Both male and female samples also contain some processes which cannot be directly related to AD.

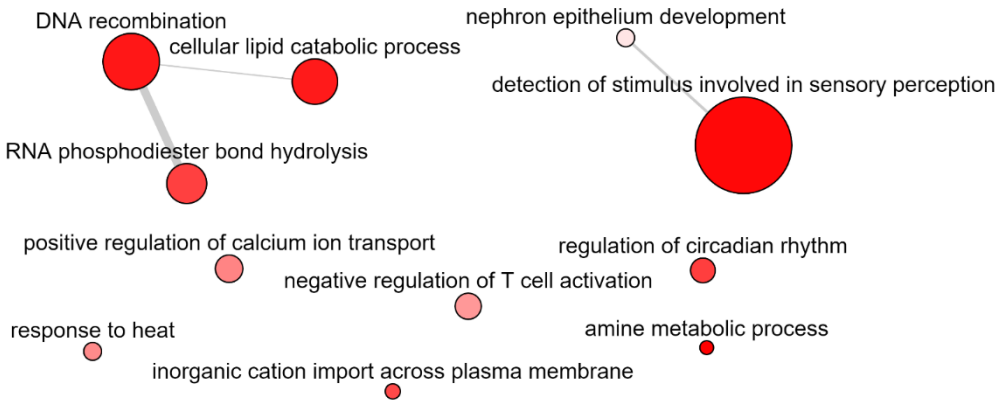

Figure 46 – Biological processes that are affected in AD, specific to the female samples. The size of the bubbles represents the number of genes related to that GO term in Homo sapiens GO database. The color scale represents the p-value: The lighter the color, the more significant the hit.

Figure 47 – Biological processes affected in AD, specific to male samples. The size of the bubbles represents the number of genes related to that GO term in the Homo sapiens GO database. The color scale represents the p-value: The lighter the color, the more significant the hit.

Most importantly, these processes significantly differ between AD and healthy samples but cannot be observed in cell models. The processes related to molecule transportation, signaling, and membrane transportation are related to cell-to-cell communication with high probability. It is expected to see these in tissue samples but not in cell models, as cell models are created in specific environments, so the extracellular environment may not support this communication, and the connection between different cell types is missing, or they only exist primarily in two dimensions, in contrast to the interactions in brain tissue.

These results also tell a story about processes related to mitotic cell cycles and phase transitions, which are also missing from the cell models. This might be caused by the forced transformation and proliferation of iPSCs to specific cells, leading to an unnatural transcriptomic profile, unrepresentative of actual tissues.

In conclusion, when working on AD with cell models, it is possible to miss processes related to mitotic cycles, cell communications, and environmental stimuli.

Neurons are the most similar cell models to tissue samples in terms of the transcriptomic profile; the similarity further increases for male samples. “Neuron – male sample” similarity is the highest similarity we observed (42,79) in tissue-cell comparisons. It is even higher than almost all tissue-tissue comparisons (Figure 32). This suggests that neurons are the best candidate for transcriptomic research in AD when cell models are considered.

4.1. Conclusion

As a summary, GSE comparisons are providing us these results: 1) If cell model studies are not enriching pathways that are related to cell-to-cell communication, DNA recombination, ion and vesicle transportation, environmental response (including response to changes in the metabolism of the organism), 2) If cell model studies are enriching pathways that are normally active during the development of the organism and the cells, this might be results of the limitations of the current methods while using cell models for AD studies. These results can be generalized for other disease models, too.

In addition, cell models can also provide novel information, normally hidden behind the background data of tissues studies, such as the ERBB pathway. It was not found in any of the tissues comparisons, while it was found in neuron and microglia models.

4.2. Future Studies

These results provide a peek to the potential of cell models in AD. Since they are built upon the available resources as of now, the results are not final, and they may never be. To improve our findings, more projects from RNA-Seq studies should be included, to begin with. In addition, studies from other fields can be merged with the RNA-Seq field, such as microRNA studies as they are interconnected, and scRNA-Seq studies, as they are gaining popularity and providing a lot of useful data.

The topology plots can be improved by positioning similar GO terms together. Currently, the order of GO terms on the plots is random, and the positions are meaningless. As GO terms are hierarchical, similar terms and ones under the same branch can be packed together, which would make it easier to compare the topologies. A “zoom” function can be added for users that want to increase the detail of the plots.

Terms from other databases can be added to topology plots, just like BP, MF, and CC from GO terms. This can allow for comparisons between the results taken from other popular databases, such as KEGG or Reactome.

Oligodendrocyte studies could also bear useful results, as they also carry an AD pathology as the disease develops, yet, our search results didn’t provide any studies regarding them in AD.

The overall workflow of the study can be made more flexible to allow for other comparisons, especially on the gene ontology step, since the tools that are used to analyze GO data are built for only a single dataset and don't consider a comparison between many samples. The comparison methods used in this study can be improved by taking the cell models that are proven to be useful as a reference. If enough data is gathered, machine learning methods can be integrated for fine-tuning the parameters when making a comparison between two novel studies. This flexibility can also bring comparisons with up and downregulated genes, as this was not considered in this study.

REFERENCES

ABCA7 ATP binding cassette subfamily A member 7 [Homo sapiens (human)] - Gene - NCBI. (n.d.). Retrieved January 8, 2023, from https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=10347

Agulhon, C., Fiacco, T. A., & McCarthy, K. D. (2010). Hippocampal short- and long-term plasticity are not modulated by astrocyte Ca2+ signaling. *Science (New York, N.Y.)*, *327*(5970), 1250–1254. https://doi.org/10.1126/SCIENCE.1184821

Allen, G., Barnard, H., McColl, R., Hester, A. L., Fields, J. A., Weiner, M. F., Ringe, W. K., Lipton, A. M., Brooker, M., McDonald, E., Rubin, C. D., & Cullum, C. M. (2007). Reduced hippocampal functional connectivity in Alzheimer disease. *Archives of Neurology*, *64*(10), 1482–1487. https://doi.org/10.1001/ARCHNEUR.64.10.1482

Atri, A. (2019). The Alzheimer's Disease Clinical Spectrum: Diagnosis and Management. *The Medical Clinics of North America*, *103*(2), 263–293. https://doi.org/10.1016/J.MCNA.2018.10.009

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. (n.d.). Retrieved January 1, 2023, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Ballard, D., Winkler-Galicki, J., & Wesoły, J. (2020). Massive parallel sequencing in forensics: advantages, issues, technicalities, and prospects. *International Journal of Legal Medicine*, *134*(4), 1291. https://doi.org/10.1007/S00414-020-02294-0

Balsis, S., Benge, J. F., Lowe, D. A., Geraci, L., & Doody, R. S. (2015). How Do Scores on the ADAS-Cog, MMSE, and CDR-SOB Correspond? *The Clinical Neuropsychologist*, *29*(7), 1002–1009. https://doi.org/10.1080/13854046.2015.1119312

Bandyopadhyay, S. (2021). Role of Neuron and Glia in Alzheimer's Disease and Associated Vascular Dysfunction. *Frontiers in Aging Neuroscience*, *13*, 274. https://doi.org/10.3389/FNAGI.2021.653334/BIBTEX

Baumann, N., & Pham-Dinh, D. (2001). Biology of oligodendrocyte and myelin in the mammalian central nervous system. *Physiological Reviews*, *81*(2), 871–927. https://doi.org/10.1152/PHYSREV.2001.81.2.871

Bioconductor - EnhancedVolcano. (n.d.). Retrieved January 17, 2023, from https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html

Bioconductor - org.Hs.eg.db. (n.d.). Retrieved January 17, 2023, from https://bioconductor.org/packages/release/data/annotation/html/org.Hs.eg.db.html

Bleidorn, C. (2015). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. *Http://Dx.Doi.Org/10.1080/14772000.2015.1099575*, *14*(1), 1–8. https://doi.org/10.1080/14772000.2015.1099575

Boedeker, S., Halstenberg, E., Schulz, P., Beblo, T., Kreisel, S., Driessen, M., & Toepper, M. (2022). Impaired Color Discrimination in Alzheimer Disease Dementia. *Alzheimer Disease and Associated Disorders*, *36*(1), 22–28. https://doi.org/10.1097/WAD.0000000000000478

Braak, H., & Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathologica*, *82*(4), 239–259. https://doi.org/10.1007/BF00308809

Branton, D., Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., di Ventra, M., Garaj, S., Hibbs, A., Huang, X., Jovanovich, S. B., Krstic, P. S., Lindsay, S., Ling, X. S., Mastrangelo, C. H., Meller, A., Oliver, J. S., Pershin, Y. v., Ramsey, J. M., … Schloss, J. A. (2008). The potential and challenges of nanopore sequencing. *Nature Biotechnology*, *26*(10), 1146–1153. https://doi.org/10.1038/NBT.1495

Brier, M. R., Thomas, J. B., Snyder, A. Z., Benzinger, T. L., Zhang, D., Raichle, M. E., Holtzman, D. M., Morris, J. C., & Ances Dr., B. M. (2012). Loss of intranetwork and internetwork resting state functional connections with Alzheimer's disease progression. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(26), 8890–8899. https://doi.org/10.1523/JNEUROSCI.5698-11.2012

Buonanno, A., & Fischbach, G. D. (2001). Neuregulin and ErbB receptor signaling pathways in the nervous system. *Current Opinion in Neurobiology*, *11*(3), 287–296. https://doi.org/10.1016/S0959-4388(00)00210-5

Burda, J. E., & Sofroniew, M. v. (2014). Reactive gliosis and the multicellular response to CNS damage and disease. *Neuron*, *81*(2), 229–248. https://doi.org/10.1016/J.NEURON.2013.12.034

Cai, Z., & Xiao, M. (2016). Oligodendrocytes and Alzheimer's disease. *The International Journal of Neuroscience*, *126*(2), 97–104. https://doi.org/10.3109/00207454.2015.1025778

Campbell, K., & Götz, M. (2002). Radial glia: multi-purpose cells for vertebrate brain development. *Trends in Neurosciences*, *25*(5), 235–238. https://doi.org/10.1016/S0166-2236(02)02156-2

Cardona, K., Medina, J., Orrego-Cardozo, M., de Mejía, F. R., Elcoroaristizabal, X., & Galvis, C. A. N. (2021). Inflammatory gene expression profiling in peripheral blood from patients with Alzheimer's disease reveals key pathways and hub genes with potential diagnostic utility: a preliminary study. *PeerJ*, *9*. https://doi.org/10.7717/PEERJ.12016/SUPP-1

Carlén, M., Meletis, K., Göritz, C., Darsalia, V., Evergren, E., Tanigaki, K., Amendola, M., Barnabé-Heider, F., Yeung, M. S. Y., Naldini, L., Honjo, T., Kokaia, Z., Shupliakov, O., Cassidy, R. M., Lindvall, O., & Frisén, J. (2009). Forebrain ependymal cells are Notch-dependent and generate neuroblasts and astrocytes after stroke. *Nature Neuroscience*, *12*(3), 259–267. https://doi.org/10.1038/NN.2268

Carr, A. C., & Moore, S. D. (2012). Robust Quantification of Polymerase Chain Reactions Using Global Fitting. *PLoS ONE*, *7*(5). https://doi.org/10.1371/JOURNAL.PONE.0037640

Cetin, S., Knez, D., Gobec, S., Kos, J., & Pišlar, A. (2022). Cell models for Alzheimer's and Parkinson's disease: At the interface of biology and drug discovery. *Biomedicine & Pharmacotherapy*, *149*, 112924. https://doi.org/10.1016/J.BIOPHA.2022.112924

Chen, X., Sun, G., Tian, E., Zhang, M., Davtyan, H., Beach, T. G., Reiman, E. M., Blurton-Jones, M., Holtzman, D. M., & Shi, Y. (2021a). Modeling Sporadic Alzheimer's Disease in Human Brain Organoids under Serum Exposure. *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)*, *8*(18). https://doi.org/10.1002/ADVS.202101462

Chen, X., Sun, G., Tian, E., Zhang, M., Davtyan, H., Beach, T. G., Reiman, E. M., Blurton-Jones, M., Holtzman, D. M., & Shi, Y. (2021b). Modeling Sporadic Alzheimer's Disease in Human Brain Organoids under Serum Exposure. *Advanced Science (Weinheim, Baden-Wurttemberg, Germany)*, *8*(18). https://doi.org/10.1002/ADVS.202101462

Cheng, S. T. (2016). Cognitive Reserve and the Prevention of Dementia: the Role of Physical and Cognitive Activities. *Current Psychiatry Reports*, *18*(9). https://doi.org/10.1007/S11920-016-0721-2

Corder, E. H., Saunders, A. M., Risch, N. J., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Rimmler, J. B., Locke, P. A., Conneally, P. M., Schmader, K. E., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1994). Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. *Nature Genetics*, *7*(2), 180–184. https://doi.org/10.1038/NG0694-180

CRAN - Package pheatmap. (n.d.). Retrieved January 17, 2023, from https://cran.r-project.org/web/packages/pheatmap/index.html

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., da Rin Fioretto, L., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. https://doi.org/10.1093/NAR/GKAB1049

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/GIGASCIENCE/GIAB008

DeMattos, R. B., Cirrito, J. R., Parsadanian, M., May, P. C., O'Dell, M. A., Taylor, J. W., Harmony, J. A. K., Aronow, B. J., Bales, K. R., Paul, S. M., & Holtzman, D. M. (2004). ApoE and Clusterin Cooperatively Suppress Aβ Levels and Deposition: Evidence that ApoE Regulates Extracellular Aβ Metabolism In Vivo. *Neuron*, *41*(2), 193–202. https://doi.org/10.1016/S0896-6273(03)00850-X

Deture, M. A., & Dickson, D. W. (2019). The neuropathological diagnosis of Alzheimer's disease. *Molecular Neurodegeneration 2019 14:1*, *14*(1), 1–18. https://doi.org/10.1186/S13024-019-0333-5

Dissing-Olesen, L., Ladeby, R., Nielsen, H. H., Toft-Hansen, H., Dalmau, I., & Finsen, B. (2007). Axonal lesion-induced microglial proliferation and microglial cluster formation in the mouse. *Neuroscience*, *149*(1), 112–122. https://doi.org/10.1016/J.NEUROSCIENCE.2007.06.037

Dolmetsch, R., & Geschwind, D. H. (2011). The human brain in a dish: The promise of iPSC-derived neurons. *Cell*, *145*(6), 831. https://doi.org/10.1016/J.CELL.2011.05.034

dos Santos, S. E., Medeiros, M., Porfirio, J., Tavares, W., Pessôa, L., Grinberg, L., Leite, R. E. P., Ferretti-Rebustini, R. E. L., Suemoto, C. K., Filho, W. J., Noctor, S. C., Sherwood, C. C., Kaas, J. H., Manger, P. R., & Herculano-Houzel, S. (2020).

Similar Microglial Cell Densities across Brain Structures and Mammalian Species: Implications for Brain Tissue Function. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *40*(24), 4622–4643. https://doi.org/10.1523/JNEUROSCI.2339-19.2020

Dourlen, P., Kilinc, D., Malmanche, N., Chapuis, J., & Lambert, J. C. (2019a). The new genetic landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? *Acta Neuropathologica 2019 138:2*, *138*(2), 221–236. https://doi.org/10.1007/S00401-019-02004-0

Dourlen, P., Kilinc, D., Malmanche, N., Chapuis, J., & Lambert, J. C. (2019b). The new genetic landscape of Alzheimer's disease: from amyloid cascade to genetically driven synaptic failure hypothesis? *Acta Neuropathologica 2019 138:2*, *138*(2), 221–236. https://doi.org/10.1007/S00401-019-02004-0

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science (New York, N.Y.)*, *323*(5910), 133–138. https://doi.org/10.1126/SCIENCE.1162986

EPHA1 EPH receptor A1 [Homo sapiens (human)] - Gene - NCBI. (n.d.). Retrieved January 8, 2023, from https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=ShowDetailView&TermToSearch=2041

FASTQ file format example. | Download Scientific Diagram. (n.d.). Retrieved January 1, 2023, from https://www.researchgate.net/figure/FASTQ-file-format-example_fig8_303293093

FastQC A Quality Control tool for High Throughput Sequence Data. (n.d.). Retrieved January 17, 2023, from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Fields, R. D., Araque, A., Johansen-Berg, H., Lim, S. S., Lynch, G., Nave, K. A., Nedergaard, M., Perez, R., Sejnowski, T., & Wake, H. (2014). Glial Biology in Learning and Cognition. *The Neuroscientist*, *20*(5), 426. https://doi.org/10.1177/1073858413504465

File:Neuron with oligodendrocyte and myelin sheath.svg - Wikimedia Commons. (n.d.). Retrieved January 13, 2023, from https://commons.wikimedia.org/wiki/File:Neuron_with_oligodendrocyte_and_myelin_sheath.svg

Filiano, A. J., Gadani, S. P., & Kipnis, J. (2015). Interactions of innate and adaptive immunity in brain development and function. *Brain Research*, *1617*, 18–27. https://doi.org/10.1016/J.BRAINRES.2014.07.050

Fink, H. A., Linskens, E. J., MacDonald, R., Silverman, P. C., McCarten, J. R., Talley, K. M. C., Forte, M. L., Desai, P. J., Nelson, V. A., Miller, M. A., Hemmy, L. S., Brasure, M., Taylor, B. C., Ng, W., Ouellette, J. M., Sheets, K. M., Wilt, T. J., & Butler, M. (2020). Benefits and Harms of Prescription Drugs and Supplements for Treatment of Clinical Alzheimer-Type Dementia. *Annals of Internal Medicine*, *172*(10), 656–668. https://doi.org/10.7326/M19-3887

Freeman, M. R., & Rowitch, D. H. (2013). Evolving Concepts of Gliogenesis: A Look Way Back and Ahead to the Next 25 Years. *Neuron*, *80*(3), 613. https://doi.org/10.1016/J.NEURON.2013.10.034

Gao, Y., Liu, J., Wang, J., Liu, Y., Zeng, L. H., Ge, W., & Ma, C. (2022). Proteomic analysis of human hippocampal subfields provides new insights into the pathogenesis of Alzheimer's disease and the role of glial cells. *Brain Pathology (Zurich, Switzerland)*, *32*(4). https://doi.org/10.1111/BPA.13047

Garnache-Ottou, F., Chaperot, L., Biichle, S., Ferrand, C., Remy-Martin, J. P., Deconinck, E., de Tailly, P. D., Bulabois, B., Poulet, J., Kuhlein, E., Jacob, M. C., Salaun, V., Arock, M., Drenou, B., Schillinger, F., Seilles, E., Tiberghien, P., Bensa, J. C., Plumas, J., & Saas, P. (2005). Expression of the myeloid-associated marker CD33 is not an exclusive factor for leukemic plasmacytoid dendritic cells. *Blood*, *105*(3), 1256–1264. https://doi.org/10.1182/BLOOD-2004-06-2416

Garofalo, M., Pandini, C., Bordoni, M., Pansarasa, O., Rey, F., Costa, A., Minafra, B., Diamanti, L., Zucca, S., Carelli, S., Cereda, C., & Gagliardi, S. (2020). Alzheimer's, Parkinson's Disease and Amyotrophic Lateral Sclerosis Gene Expression Patterns Divergence Reveals Different Grade of RNA Metabolism Involvement. *International Journal of Molecular Sciences*, *21*(24), 1–16. https://doi.org/10.3390/IJMS21249500

Gatz, M., Reynolds, C. A., Fratiglioni, L., Johansson, B., Mortimer, J. A., Berg, S., Fiske, A., & Pedersen, N. L. (2006). Role of genes and environments for explaining Alzheimer disease. *Archives of General Psychiatry*, *63*(2), 168–174. https://doi.org/10.1001/ARCHPSYC.63.2.168

Gehrmann, J., Matsumoto, Y., & Kreutzberg, G. W. (1995). Microglia: Intrinsic immuneffector cell of the brain. *Brain Research Reviews*, *20*(3), 269–287. https://doi.org/10.1016/0165-0173(94)00015-H

Ginestet, C. (2011). ggplot2: Elegant Graphics for Data Analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *174*(1). https://doi.org/10.1111/j.1467-985x.2010.00676_9.x

Ginhoux, F., Lim, S., Hoeffel, G., Low, D., & Huber, T. (2013). Origin and differentiation of microglia. *Frontiers in Cellular Neuroscience*, *7*(MAR). https://doi.org/10.3389/FNCEL.2013.00045

Goate, A., Chartier-Harlin, M. C., Mullan, M., Brown, J., Crawford, F., Fidani, L., Giuffra, L., Haynes, A., Irving, N., James, L., Mant, R., Newton, P., Rooke, K., Roques, P., Talbot, C., Pericak-Vance, M., Roses, A., Williamson, R., Rossor, M., … Hardy, J. (1991). Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease. *Nature*, *349*(6311), 704–706. https://doi.org/10.1038/349704A0

Goedert, M., & Spillantini, M. G. (2006). A century of Alzheimer's disease. *Science (New York, N.Y.)*, *314*(5800), 777–781. https://doi.org/10.1126/SCIENCE.1132814

Gomez, W., Morales, R., Maracaja-Coutinho, V., Parra, V., & Nassif, M. (2020). Down syndrome and Alzheimer's disease: common molecular traits beyond the amyloid precursor protein. *Aging*, *12*(1), 1011–1033. https://doi.org/10.18632/AGING.102677

Gomez-Nicola, D., & Boche, D. (2015). Post-mortem analysis of neuroinflammatory changes in human Alzheimer's disease. *Alzheimer's Research & Therapy*, *7*(1). https://doi.org/10.1186/S13195-015-0126-1

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics 2016 17:6*, *17*(6), 333–351. https://doi.org/10.1038/nrg.2016.49

Grossberg, G. T. (2002). The ABC of Alzheimer's Disease: Behavioral Symptoms and Their Treatment. *International Psychogeriatrics*, *14*(S1), 27–49. https://doi.org/10.1017/S1041610203008652

Guan, Y.-H., Zhang, L.-J., Wang, S.-Y., Deng, Y.-D., Zhou, H.-S., Chen, D.-Q., Zhang, L.-C., & Lan-Chun Zhang, C. (2022). The role of microglia in Alzheimer's disease and progress of treatment. *Ibrain*, *8*(1), 37–47. https://doi.org/10.1002/IBRA.12023

Gupta, P. K. (2008). Single-molecule DNA sequencing technologies for future genomics research. *Trends in Biotechnology*, *26*(11), 602–611. https://doi.org/10.1016/J.TIBTECH.2008.07.003

Hanna, R., Flamier, A., Barabino, A., & Bernier, G. (2021). G-quadruplexes originating from evolutionary conserved L1 elements interfere with neuronal gene expression in Alzheimer's disease. *Nature Communications*, *12*(1). https://doi.org/10.1038/S41467-021-22129-9

Hansen, D. v., Hanson, J. E., & Sheng, M. (2018). Microglia in Alzheimer's disease. *The Journal of Cell Biology*, *217*(2), 459. https://doi.org/10.1083/JCB.201709069

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A. R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M. K., Brayne, C., … Williams, J. (2009). Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nature Genetics*, *41*(10), 1088–1093. https://doi.org/10.1038/NG.440

Harris, M. A., Deegan, J. I., Ireland, A., Lomax, J., Ashburner, M., Tweedie, S., Carbon, S., Lewis, S., Mungall, C., Day-Richter, J., Eilbeck, K., Blake, J. A., Bult, C., Diehl, A. D., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., … McCarthy, F. (2008). The Gene Ontology project in 2008. *Nucleic Acids Research*, *36*(Database issue), D440. https://doi.org/10.1093/NAR/GKM883

Hebert, L. E., Scherr, P. A., Bienias, J. L., Bennett, D. A., & Evans, D. A. (2003). Alzheimer Disease in the US Population: Prevalence Estimates Using the 2000 Census. *Archives of Neurology*, *60*(8), 1119–1122. https://doi.org/10.1001/ARCHNEUR.60.8.1119

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, & P.E. Bourne. (2021). *RCSB PDB*. RCSB Protein Data Bank: Powerful New Tools for Exploring 3D Structures of Biological Macromolecules for Basic and Applied Research and Education in Fundamental Biology, Biomedicine, Biotechnology, Bioengineering and Energy Sciences. https://www.rcsb.org/

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., Abraham, R., Hamshere, M. L., Pahwa, J. S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., … Williams, J. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nature Genetics*, *43*(5), 429–436. https://doi.org/10.1038/NG.803

How HiFi sequencing works - PacBio. (n.d.). Retrieved January 7, 2023, from https://www.pacb.com/technology/hifi-sequencing/how-it-works/

How to interpret this duplicate sequence plot from FastQC. (n.d.). Retrieved January 3, 2023, from https://www.biostars.org/p/274259/

Integrative Genomics Viewer (IGV) tutorial - Bioinformatics Team (BioITeam) at the University of Texas - UT Austin Wikis. (n.d.). Retrieved January 1, 2023, from https://wikis.utexas.edu/display/bioiteam/Integrative+Genomics+Viewer+%28IGV%29+tutorial

Ivashko-Pachima, Y., Hadar, A., Grigg, I., Korenková, V., Kapitansky, O., Karmon, G., Gershovits, M., Sayas, C. L., Kooy, R. F., Attems, J., Gurwitz, D., & Gozes, I. (2021). Discovery of autism/intellectual disability somatic mutations in Alzheimer's brains: mutated ADNP cytoskeletal impairments and repair as a case study. *Molecular Psychiatry*, *26*(5), 1619–1633. https://doi.org/10.1038/S41380-019-0563-5

Jiang, Q., Lee, C. Y. D., Mandrekar, S., Wilkinson, B., Cramer, P., Zelcer, N., Mann, K., Lamb, B., Willson, T. M., Collins, J. L., Richardson, J. C., Smith, J. D., Comery, T. A., Riddell, D., Holtzman, D. M., Tontonoz, P., & Landreth, G. E. (2008). ApoE promotes the proteolytic degradation of Abeta. *Neuron*, *58*(5), 681–693. https://doi.org/10.1016/J.NEURON.2008.04.010

Jones, S. E., & Jomary, C. (2002). Clusterin. *International Journal of Biochemistry and Cell Biology*, *34*(5), 427–431. https://doi.org/10.1016/S1357-2725(01)00155-8

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology 2019 37:8*, *37*(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Kim, E., Lee, Y., Choi, S., & Song, J. J. (2014). Structural basis of the phosphorylation dependent complex formation of neurodegenerative disease protein Ataxin-1 and RBM17. *Biochemical and Biophysical Research Communications*, *449*(4), 399–404. https://doi.org/10.1016/J.BBRC.2014.05.063

Konttinen, H., Cabral-da-Silva, M. e. C., Ohtonen, S., Wojciechowski, S., Shakirzyanova, A., Caligola, S., Giugno, R., Ishchenko, Y., Hernández, D., Fazaludeen, M. F., Eamen, S., Budia, M. G., Fagerlund, I., Scoyni, F., Korhonen, P., Huber, N., Haapasalo, A., Hewitt, A. W., Vickers, J., … Malm, T. (2019). PSEN1ΔE9, APPswe, and APOE4 Confer Disparate Phenotypes in Human iPSC-Derived Microglia. *Stem Cell Reports*, *13*(4), 669–683. https://doi.org/10.1016/J.STEMCR.2019.08.004

Kwart, D., Gregg, A., Scheckel, C., Murphy, E., Paquet, D., Duffield, M., Fak, J., Olsen, O., Darnell, R., & Tessier-Lavigne, M. (2019). A Large Panel of Isogenic APP and PSEN1 Mutant Human iPSC Neurons Reveals Shared Endosomal Abnormalities

Mediated by APP β-CTFs, Not Aβ. *Neuron*, *104*(2), 256-270.e5. https://doi.org/10.1016/J.NEURON.2019.07.010

Lambert, J. C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M. J., Tavernier, B., Letenneur, L., Bettens, K., Berr, C., Pasquier, F., Fiévet, N., Barberger-Gateau, P., Engelborghs, S., de Deyn, P., Mateo, I., … Amouyel, P. (2009). Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. *Nature Genetics*, *41*(10), 1094–1099. https://doi.org/10.1038/NG.439

Lanoiselée, H. M., Nicolas, G., Wallon, D., Rovelet-Lecrux, A., Lacour, M., Rousseau, S., Richard, A. C., Pasquier, F., Rollin-Sillaire, A., Martinaud, O., Quillard-Muraine, M., de la Sayette, V., Boutoleau-Bretonniere, C., Etcharry-Bouyx, F., Chauviré, V., Sarazin, M., le Ber, I., Epelbaum, S., Jonveaux, T., … Campion, D. (2017). APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases. *PLoS Medicine*, *14*(3). https://doi.org/10.1371/JOURNAL.PMED.1002270

Lefterov, I., Wolfe, C. M., Fitz, N. F., Nam, K. N., Letronne, F., Biedrzycki, R. J., Kofler, J., Han, X., Wang, J., Schug, J., & Koldamova, R. (2019). APOE2 orchestrated differences in transcriptomic and lipidomic profiles of postmortem AD brain. *Alzheimer's Research & Therapy*, *11*(1). https://doi.org/10.1186/S13195-019-0558-0

Leng, F., & Edison, P. (2020). Neuroinflammation and microglial activation in Alzheimer disease: where do we go from here? *Nature Reviews Neurology 2020 17:3*, *17*(3), 157–172. https://doi.org/10.1038/s41582-020-00435-y

Li, K., Li, J., Zheng, J., & Qin, S. (2019). Reactive Astrocytes in Neurodegenerative Diseases. *Aging and Disease*, *10*(3), 664–675. https://doi.org/10.14336/AD.2018.0720

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/BIOINFORMATICS/BTT656

Liaoi, Y. F., Wang, B. J., Cheng, H. T., Kuo, L. H., & Wolfe, M. S. (2004). Tumor necrosis factor-alpha, interleukin-1beta, and interferon-gamma stimulate gamma-secretase-mediated cleavage of amyloid precursor protein through a JNK-dependent MAPK pathway. *The Journal of Biological Chemistry*, *279*(47), 49523–49532. https://doi.org/10.1074/JBC.M402034200

Lin, Y. T., Seo, J., Gao, F., Feldman, H. M., Wen, H. L., Penney, J., Cam, H. P., Gjoneska, E., Raja, W. K., Cheng, J., Rueda, R., Kritskiy, O., Abdurrob, F., Peng, Z., Milo, B., Yu, C. J., Elmsaouri, S., Dey, D., Ko, T., … Tsai, L. H. (2018). APOE4

Causes Widespread Molecular and Cellular Alterations Associated with Alzheimer's Disease Phenotypes in Human iPSC-Derived Brain Cell Types. *Neuron*, *98*(6), 1141-1154.e7. https://doi.org/10.1016/J.NEURON.2018.05.008

Liu, C. C., Kanekiyo, T., Xu, H., & Bu, G. (2013). Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature Reviews. Neurology*, *9*(2), 106–118. https://doi.org/10.1038/NRNEUROL.2012.263

Liu, T., Zhu, B., Liu, Y., Zhang, X., Yin, J., Li, X., Jiang, L. L., Hodges, A. P., Rosenthal, S. B., Zhou, L., Yancey, J., McQuade, A., Blurton-Jones, M., Tanzi, R. E., Huang, T. Y., & Xu, H. (2020a). Multi-omic comparison of Alzheimer's variants in human ESC-derived microglia reveals convergence at APOE. *The Journal of Experimental Medicine*, *217*(12). https://doi.org/10.1084/JEM.20200474

Liu, T., Zhu, B., Liu, Y., Zhang, X., Yin, J., Li, X., Jiang, L. L., Hodges, A. P., Rosenthal, S. B., Zhou, L., Yancey, J., McQuade, A., Blurton-Jones, M., Tanzi, R. E., Huang, T. Y., & Xu, H. (2020b). Multi-omic comparison of Alzheimer's variants in human ESC-derived microglia reveals convergence at APOE. *The Journal of Experimental Medicine*, *217*(12). https://doi.org/10.1084/JEM.20200474

Loman, N. J., & Watson, M. (2015). Successful test launch for nanopore sequencing. *Nature Methods*, *12*(4), 303–304. https://doi.org/10.1038/NMETH.3327

Long, J. M., & Holtzman, D. M. (2019). Alzheimer Disease: An Update on Pathobiology and Treatment Strategies. *Cell*, *179*(2), 312–339. https://doi.org/10.1016/J.CELL.2019.09.001

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

MacDonald, A., Lu, B., Caron, M., Caporicci-Dinucci, N., Hatrock, D., Petrecca, K., Bourque, G., & Stratton, J. A. (2021). Single Cell Transcriptomics of Ependymal Cells Across Age, Region and Species Reveals Cilia-Related and Metal Ion Regulatory Roles as Major Conserved Ependymal Cell Functions. *Frontiers in Cellular Neuroscience*, *15*, 268. https://doi.org/10.3389/FNCEL.2021.703951/BIBTEX

Magistri, M., Velmeshev, D., Makhmutova, M., & Faghihi, M. A. (2015). Transcriptomics Profiling of Alzheimer's Disease Reveal Neurovascular Defects, Altered Amyloid-β Homeostasis, and Deregulated Expression of Long Noncoding RNAs. *Journal of Alzheimer's Disease : JAD*, *48*(3), 647–665. https://doi.org/10.3233/JAD-150398

Marlatt, M. W., Bauer, J., Aronica, E., van Haastert, E. S., Hoozemans, J. J. M., Joels, M., & Lucassen, P. J. (2014). Proliferation in the Alzheimer hippocampus is due

to microglia, not astroglia, and occurs at sites of amyloid deposition. *Neural Plasticity*, *2014*. https://doi.org/10.1155/2014/693851

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/EJ.17.1.200

Matos, B., Publicover, S. J., Castro, L. F. C., Esteves, P. J., & Fardilha, M. (2021). Brain and testis: more alike than previously thought?. Open biology, 11(6), 200322. https://doi.org/10.1098/rsob.200322

Mayeux, R., & Stern, Y. (2012). Epidemiology of Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(8), a006239. https://doi.org/10.1101/CSHPERSPECT.A006239

McDermott, K. W., Barry, D. S., & McMahon, S. S. (2005). Role of radial glia in cytogenesis, patterning and boundary formation in the developing spinal cord. *Journal of Anatomy*, *207*(3), 241–250. https://doi.org/10.1111/J.1469-7580.2005.00462.X

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269. https://doi.org/10.1016/J.JALZ.2011.03.005

McQuade, A., Kang, Y. J., Hasselmann, J., Jairaman, A., Sotelo, A., Coburn, M., Shabestari, S. K., Chadarevian, J. P., Fote, G., Tu, C. H., Danhash, E., Silva, J., Martinez, E., Cotman, C., Prieto, G. A., Thompson, L. M., Steffan, J. S., Smith, I., Davtyan, H., … Blurton-Jones, M. (2020). Gene expression and functional deficits underlie TREM2-knockout microglia responses in human models of Alzheimer's disease. *Nature Communications*, *11*(1). https://doi.org/10.1038/S41467-020-19227-5

Meyer, K., Feldman, H. M., Lu, T., Drake, D., Lim, E. T., Ling, K. H., Bishop, N. A., Pan, Y., Seo, J., Lin, Y. T., Su, S. C., Church, G. M., Tsai, L. H., & Yankner, B. A. (2019). REST and Neural Gene Network Dysregulation in iPSC Models of Alzheimer's Disease. *Cell Reports*, *26*(5), 1112-1127.e9. https://doi.org/10.1016/J.CELREP.2019.01.023

Mizuno, Y., Abolhassani, N., Mazzei, G., Sakumi, K., Saito, T., Saido, T. C., Ninomiya, T., Iwaki, T., Yamasaki, R., Kira, J. I., & Nakabeppu, Y. (2021). MUTYH Actively

Contributes to Microglial Activation and Impaired Neurogenesis in the Pathogenesis of Alzheimer's Disease. *Oxidative Medicine and Cellular Longevity*, *2021*. https://doi.org/10.1155/2021/8635088

Mohs, R. C. (1996). The Alzheimer's Disease Assessment Scale. *International Psychogeriatrics*, *8*(2), 195–203. https://doi.org/10.1017/S1041610296002578

Monterey, M. D., Wei, H., Wu, X., & Wu, J. Q. (2021). The Many Faces of Astrocytes in Alzheimer's Disease. *Frontiers in Neurology*, *12*, 1402. https://doi.org/10.3389/FNEUR.2021.619626/BIBTEX

Morris, J. C. (1997). Clinical Dementia Rating: A Reliable and Valid Diagnostic and Staging Measure for Dementia of the Alzheimer Type. *International Psychogeriatrics*, *9*(S1), 173–176. https://doi.org/10.1017/S1041610297004870

Naresh, A., Long, W., Vidal, G. A., Wimley, W. C., Marrero, L., Sartor, C. I., Tovey, S., Cooke, T. G., Bartlett, J. M. S., & Jones, F. E. (2006). The ERBB4/HER4 intracellular domain 4ICD is a BH3-only protein promoting apoptosis of breast cancer cells. *Cancer Research*, *66*(12), 6412–6420. https://doi.org/10.1158/0008-5472.CAN-05-2368

Nativio, R., Lan, Y., Donahue, G., Sidoli, S., Berson, A., Srinivasan, A. R., Shcherbakova, O., Amlie-Wolf, A., Nie, J., Cui, X., He, C., Wang, L. S., Garcia, B. A., Trojanowski, J. Q., Bonini, N. M., & Berger, S. L. (2020). An integrated multi-omics approach identifies epigenetic alterations associated with Alzheimer's disease. *Nature Genetics*, *52*(10), 1024–1035. https://doi.org/10.1038/S41588-020-0696-0

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & Stamatoyannopoulos, J. A. (2012). BEDOPS: high-performance genomic feature operations. *Bioinformatics*, *28*(14), 1919–1920. https://doi.org/10.1093/BIOINFORMATICS/BTS277

Nuutinen, T., Suuronen, T., Kauppinen, A., & Salminen, A. (2009). Clusterin: a forgotten player in Alzheimer's disease. *Brain Research Reviews*, *61*(2), 89–104. https://doi.org/10.1016/J.BRAINRESREV.2009.05.007

Oksanen, M., Hyötyläinen, I., Trontti, K., Rolova, T., Wojciechowski, S., Koskuvi, M., Viitanen, M., Levonen, A. L., Hovatta, I., Roybon, L., Lehtonen, Š., Kanninen, K. M., Hämäläinen, R. H., & Koistinaho, J. (2020). NF-E2-related factor 2 activation boosts antioxidant defenses and ameliorates inflammatory and amyloid properties in human Presenilin-1 mutated Alzheimer's disease astrocytes. *Glia*, *68*(3), 589–599. https://doi.org/10.1002/GLIA.23741

Perl, D. P. (2010). Neuropathology of Alzheimer's Disease. *The Mount Sinai Journal of Medicine, New York*, *77*(1), 32. https://doi.org/10.1002/MSJ.20157

Piguet, O., Double, K. L., Kril, J. J., Harasty, J., Macdonald, V., McRitchie, D. A., & Halliday, G. M. (2009). White matter loss in healthy ageing: a postmortem analysis. *Neurobiology of Aging*, *30*(8), 1288–1295. https://doi.org/10.1016/J.NEUROBIOLAGING.2007.10.015

Potemkin, N., Cawood, S. M. F., Treece, J., Guévremont, D., Rand, C. J., McLean, C., Stanton, J. A. L., & Williams, J. M. (2022). A method for simultaneous detection of small and long RNA biotypes by ribodepleted RNA-Seq. *Scientific Reports*, *12*(1). https://doi.org/10.1038/S41598-021-04209-4

Priller, C., Bauer, T., Mitteregger, G., Krebs, B., Kretzschmar, H. A., & Herms, J. (2006). Synapse formation and function is modulated by the amyloid precursor protein. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(27), 7212–7221. https://doi.org/10.1523/JNEUROSCI.1450-06.2006

Purves, D., Augustine, G. J., & Fitzpatrick, D. (2012). Neuroscience, 5th Edition. In *Nature Reviews Neuroscience*.

Qiang, W., Yau, W. M., Luo, Y., Mattson, M. P., & Tycko, R. (2012). Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(12), 4443–4448. https://doi.org/10.1073/PNAS.1111305109/-/DCSUPPLEMENTAL/PNAS.1111305109_SI.PDF

Redman, M., King, A., Watson, C., & King, D. (2016). What is CRISPR/Cas9? *Archives of Disease in Childhood. Education and Practice Edition*, *101*(4), 213. https://doi.org/10.1136/ARCHDISCHILD-2016-310459

Richardson, W. D., Kessaris, N., & Pringle, N. (2006). Oligodendrocyte wars. *Nature Reviews. Neuroscience*, *7*(1), 11–18. https://doi.org/10.1038/NRN1826

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463. https://doi.org/10.1073/PNAS.74.12.5463

Sardi, S. P., Murtie, J., Koirala, S., Patten, B. A., & Corfas, G. (2006). Presenilin-dependent ErbB4 nuclear signaling regulates the timing of astrogenesis in the developing brain. *Cell*, *127*(1), 185–197. https://doi.org/10.1016/J.CELL.2006.07.037

Sartori, M., Mendes, T., Desai, S., Lasorsa, A., Herledan, A., Malmanche, N., Mäkinen, P., Marttinen, M., Malki, I., Chapuis, J., Flaig, A., Vreulx, A.-C., Amouyel, P., Leroux, F., Déprez, B., Cantrelle, F.-X., Maréchal, D., Pradier, L., Hiltunen, M., … Lambert, J.-C. (2018). BIN1 recovers tauopathy-induced long-term memory deficits in mice and interacts with Tau through Thr348 phosphorylation. *BioRxiv*, *5*, 462317. https://doi.org/10.1101/462317

Satpute-Krishnan, P., DeGiorgis, J. A., Conley, M. P., Jang, M., & Bearer, E. L. (2006). A peptide zipcode sufficient for anterograde transport within amyloid precursor protein. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(44), 16532–16537. https://doi.org/10.1073/PNAS.0607527103

Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., Connor, R., Funk, K., Kelly, C., Kim, S., Madej, T., Marchler-Bauer, A., Lanczycki, C., Lathrop, S., Lu, Z., Thibaud-Nissen, F., Murphy, T., Phan, L., Skripchenko, Y., … Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Research*, *50*(D1), D20–D26. https://doi.org/10.1093/NAR/GKAB1112

Scheckel, C., Drapeau, E., Frias, M. A., Park, C. Y., Fak, J., Zucker-Scharff, I., Kou, Y., Haroutunian, V., Ma'ayan, A., Buxbaum, J. D., & Darnell, R. B. (2016). Regulatory consequences of neuronal ELAV-like protein binding to coding and non-coding RNAs in human brain. *ELife*, *5*(FEBRUARY2016). https://doi.org/10.7554/ELIFE.10421

Scheltens, P., Blennow, K., Breteler, M. M. B., de Strooper, B., Frisoni, G. B., Salloway, S., & van der Flier, W. M. (2016). Alzheimer's disease. *Lancet (London, England)*, *388*(10043), 505–517. https://doi.org/10.1016/S0140-6736(15)01124-1

Scheyer, O., Rahman, A., Hristov, H., Berkowitz, C., Isaacson, R. S., Diaz Brinton, R., & Mosconi, L. (2018). Female Sex and Alzheimer's Risk: The Menopause Connection. *The Journal of Prevention of Alzheimer's Disease*, *5*(4), 225–230. https://doi.org/10.14283/JPAD.2018.34

Sepehrnia, B., Kamboh, M. I., Adams-Campbell, L. L., Bunker, C. H., Nwankwo, M., Majumder, P. P., & Ferrell, R. E. (1989). Genetic studies of human apolipoproteins. X. The effect of the apolipoprotein E polymorphism on quantitative levels of lipoproteins in Nigerian blacks. *American Journal of Human Genetics*, *45*(4), 586. /pmc/articles/PMC1683508/?report=abstract

Serrano-Pozo, A., Frosch, M. P., Masliah, E., & Hyman, B. T. (2011). Neuropathological alterations in Alzheimer disease. *Cold Spring Harbor Perspectives in Medicine*, *1*(1). https://doi.org/10.1101/CSHPERSPECT.A006189

Sharkey, D. J., Scalice, E. R., Christy, K. G., Atwood, S. M., & Daiss, J. L. (1994). Antibodies as thermolabile switches: high temperature triggering for the polymerase chain reaction. *Bio/Technology (Nature Publishing Company)*, *12*(5), 506–509. https://doi.org/10.1038/NBT0594-506

Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. *Nature Biotechnology*, *26*(10), 1135–1145. https://doi.org/10.1038/NBT1486

Squire, L. R., Berg, D., Bloom, F. E., du Lac, S., Ghosh, A., & Spitzer, N. C. (2012). Fundamental Neuroscience: Fourth Edition. In *Fundamental Neuroscience: Fourth Edition*. https://doi.org/10.1016/C2010-0-65035-8

SRA Toolkit . (n.d.). Retrieved January 17, 2023, from https://github.com/ncbi/sra-tools/wiki/01.-Downloading-SRA-Toolkit

Staugaitis, S. M. (2011). Practical Surgical Neuropathology. *American Journal of Surgical Pathology*, *35*(1). https://doi.org/10.1097/pas.0b013e3181fa2372

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE*, *6*(7), e21800. https://doi.org/10.1371/JOURNAL.PONE.0021800

Tanzi, R. E. (2012). The Genetics of Alzheimer Disease. *Cold Spring Harbor Perspectives in Medicine*, *2*(10). https://doi.org/10.1101/CSHPERSPECT.A006296

Terwel, D., Steffensen, K. R., Verghese, P. B., Kummer, M. P., Gustafsson, J. Å., Holtzman, D. M., & Heneka, M. T. (2011). Critical role of astroglial apolipoprotein E and liver X receptor-α expression for microglial Aβ phagocytosis. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(19), 7049–7059. https://doi.org/10.1523/JNEUROSCI.6546-10.2011

Thal, D. R., Rüb, U., Orantes, M., & Braak, H. (2002). Phases of A beta-deposition in the human brain and its relevance for the development of AD. *Neurology*, *58*(12), 1791–1800. https://doi.org/10.1212/WNL.58.12.1791

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini-Mental State Examination: A Comprehensive Review. *Journal of the American Geriatrics Society*, *40*(9), 922–935. https://doi.org/10.1111/J.1532-5415.1992.TB01992.X

Tuimala, J., & Kallio, A. (2013). R, Programming Language. *Encyclopedia of Systems Biology*, 1809–1811. https://doi.org/10.1007/978-1-4419-9863-7_619

van Dyck, C. H., Swanson, C. J., Aisen, P., Bateman, R. J., Chen, C., Gee, M., Kanekiyo, M., Li, D., Reyderman, L., Cohen, S., Froelich, L., Katayama, S., Sabbagh, M., Vellas, B., Watson, D., Dhadda, S., Irizarry, M., Kramer, L. D., & Iwatsubo, T. (2022). Lecanemab in Early Alzheimer's Disease. *The New England Journal of Medicine*. https://doi.org/10.1056/NEJMOA2212948/SUPPL_FILE/NEJMOA2212948_APPENDIX.PDF

Verkhratsky, A., & Butt, A. (2013). Glial Physiology and Pathophysiology. In *Glial Physiology and Pathophysiology*. https://doi.org/10.1002/9781118402061

Wainaina, M. N., Chen, Z., & Zhong, C. (2014). Environmental factors in the development and progression of late-onset Alzheimer's disease. *Neuroscience Bulletin 2014 30:2*, *30*(2), 253–270. https://doi.org/10.1007/S12264-013-1425-9

Wang, J., Gu, B. J., Masters, C. L., & Wang, Y. J. (2017). A systemic view of Alzheimer disease - insights from amyloid-β metabolism beyond the brain. *Nature Reviews. Neurology*, *13*(10), 612–623. https://doi.org/10.1038/NRNEUROL.2017.111

Wang, L., Wang, S., & Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics (Oxford, England)*, *28*(16), 2184–2185. https://doi.org/10.1093/BIOINFORMATICS/BTS356

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology 2021 39:11*, *39*(11), 1348–1365. https://doi.org/10.1038/s41587-021-01108-x

Weiner, M. F., Hynan, L. S., Bret, M. E., & White, C. (2005). Early behavioral symptoms and course of Alzheimer's disease. *Acta Psychiatrica Scandinavica*, *111*(5), 367–371. https://doi.org/10.1111/J.1600-0447.2004.00472.X

Wightman, D. P., Jansen, I. E., Savage, J. E., Shadrin, A. A., Bahrami, S., Holland, D., Rongve, A., Børte, S., Winsvold, B. S., Drange, O. K., Martinsen, A. E., Skogholt, A. H., Willer, C., Bråthen, G., Bosnes, I., Nielsen, J. B., Fritsche, L. G., Thomas, L. F., Pedersen, L. M., … Posthuma, D. (2021). A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nature Genetics*, *53*(9), 1276–1282. https://doi.org/10.1038/S41588-021-00921-Z

Wikipedia. (2022). *Sanger sequencing*. https://en.wikipedia.org/wiki/Sanger_sequencing

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation*, *2*(3), 100141. https://doi.org/10.1016/j.xinn.2021.100141

Xiao, Y., Ma, B., McElheny, D., Parthasarathy, S., Long, F., Hoshi, M., Nussinov, R., & Ishii, Y. (2015). Aβ(1-42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Structural & Molecular Biology*, *22*(6), 499–505. https://doi.org/10.1038/NSMB.2991

Yang, H., Oh, C. K., Amal, H., Wishnok, J. S., Lewis, S., Schahrer, E., Trudler, D., Nakamura, T., Tannenbaum, S. R., & Lipton, S. A. (2022). Mechanistic insight into female predominance in Alzheimer's disease based on aberrant protein Snitrosylation of C3. *Science Advances*, *8*(50). https://doi.org/10.1126/SCIADV.ADE0764/SUPPL_FILE/SCIADV.ADE0764_TABLES_S3_TO_S14.ZIP

Yilmaz, A., Onen, H. I., Alp, E., Menevse, S., Yilmaz, A., Onen, H. I., Alp, E., & Menevse, S. (2012). Real-Time PCR for Gene Expression Analysis. *Polymerase Chain Reaction*. https://doi.org/10.5772/37356

Yu, J. T., Xu, W., Tan, C. C., Andrieu, S., Suckling, J., Evangelou, E., Pan, A., Zhang, C., Jia, J., Feng, L., Kua, E. H., Wang, Y. J., Wang, H. F., Tan, M. S., Li, J. Q., Hou, X. H., Wan, Y., Tan, L., Mok, V., … Vellas, B. (2020). Evidence-based prevention of Alzheimer's disease: systematic review and meta-analysis of 243 observational prospective studies and 153 randomised controlled trials. *Journal of Neurology, Neurosurgery & Psychiatry*, *91*(11), 1201–1209. https://doi.org/10.1136/JNNP-2019-321913

Zhang, C., Browne, A., Child, D., DiVito, J. R., Stevenson, J. A., & Tanzi, R. E. (2010). Loss of function of ATXN1 increases amyloid beta-protein levels by potentiating beta-secretase processing of beta-amyloid precursor protein. *The Journal of Biological Chemistry*, *285*(12), 8515–8526. https://doi.org/10.1074/JBC.M109.079079

Zhang, Z., Mu, J., Li, J., Li, W., & Song, J. (2013). Aberrant apolipoprotein E expression and cognitive dysfunction in patients with poststroke depression. *Genetic Testing and Molecular Biomarkers*, *17*(1), 47–51. https://doi.org/10.1089/GTMB.2012.0253

Zhu, W. Z., Xie, Y., Moyes, K. W., Gold, J. D., Askari, B., & Laflamme, M. A. (2010). Neuregulin/ErbB Signaling Regulates Cardiac Subtype Specification in Differentiating Human Embryonic Stem Cells. *Circulation Research*, *107*(6), 776. https://doi.org/10.1161/CIRCRESAHA.110.223917

APPENDICES

APPENDIX A

META INFORMATION AND PRE-ANALYSIS OF THE PROJECTS

7.1. Experimental Designs

The following projects are taken from NCBI, SRA. Details of the experiments and the samples can be found under the corresponding title below.

7.1.1. PRJNA232669

Meta information: Illumina HiSeq 2500, dorsolateral prefrontal Cortex

Overall design: Human brain samples were obtained from the Mount Sinai Brain Bank; RNA was Trizol extracted, Ribominus selected and submitted for high-throughput sequencing.

Table 9 – Design matrix of PRJNA232669

	tissue	condition	run	type	sample
SRR2422918	dorsolateral prefrontal cortex	healthy	SRR2422918	paired-end	H1
SRR2422919	dorsolateral prefrontal cortex	healthy	SRR2422919	paired-end	H2
SRR2422920	dorsolateral prefrontal cortex	healthy	SRR2422920	paired-end	H3
SRR2422921	dorsolateral prefrontal cortex	healthy	SRR2422921	paired-end	H4
SRR2422922	dorsolateral prefrontal cortex	healthy	SRR2422922	paired-end	H5
SRR2422923	dorsolateral prefrontal cortex	healthy	SRR2422923	paired-end	H6
SRR2422924	dorsolateral prefrontal cortex	healthy	SRR2422924	paired-end	H7
SRR2422925	dorsolateral prefrontal cortex	healthy	SRR2422925	paired-end	H8
SRR2422926	dorsolateral prefrontal cortex	AD	SRR2422926	paired-end	AD1
SRR2422927	dorsolateral prefrontal cortex	AD	SRR2422927	paired-end	AD2
SRR2422928	dorsolateral prefrontal cortex	AD	SRR2422928	paired-end	AD3
SRR2422929	dorsolateral prefrontal cortex	AD	SRR2422929	paired-end	AD4
SRR2422930	dorsolateral prefrontal cortex	AD	SRR2422930	paired-end	AD5
SRR2422931	dorsolateral prefrontal cortex	AD	SRR2422931	paired-end	AD6
SRR2422932	dorsolateral prefrontal cortex	AD	SRR2422932	paired-end	AD7

SRR2422933	dorsolateral prefrontal cortex	AD	SRR2422933	paired-end	AD8
SRR2422934	dorsolateral prefrontal cortex	AD	SRR2422934	paired-end	AD9

7.1.2. PRJNA279526

Meta information: Illumina HiSeq 2000, hippocampus

Overall design: Directional RNA sequencing was performed on high-quality RNA samples that were extracted from hippocampi of 4 late-onset Alzheimer's disease (LOAD) and 4 age-matched controls .

Table 10 – Design matrix of PRJNA279526

	condition	sex	APOE	braak_stage	run	type	sample
SRR1931812	LOAD	female	3-3	4	SRR1931812	single-end	AD1F
SRR1931813	LOAD	female	3-3	5	SRR1931813	single-end	AD2F
SRR1931814	LOAD	male	3-3	6	SRR1931814	single-end	AD3M
SRR1931815	LOAD	female	3-3	6	SRR1931815	single-end	AD4F
SRR1931816	healthy	female	2-3	1	SRR1931816	single-end	H1F
SRR1931817	healthy	male	NA	2	SRR1931817	single-end	H2M
SRR1931818	healthy	male	3-3	2	SRR1931818	single-end	H3M
SRR1931819	healthy	female	2-3	2	SRR1931819	single-end	H4F

7.1.3. PRJNA399530

Meta information: Illumina HiSeq 2000, iPSC-derived brain cell types

Overall design: CRISPR/Cas9 was used to create isogenic APOE3 and APOE4 human iPSC lines that enables to carry out a comprehensive analysis of biochemical, cellular, and transcriptional changes by APOE variants in multiple iPSC-derived brain cell types

Table 11 - Design matrix of PRJNA399530

	cell	APOE	Cell-line	sample	run	type
SRR5956447	iPSC	E3	32	iPSC01APOE3	SRR5956447	single-end
SRR5956448	iPSC	E3	33	iPSC02APOE3	SRR5956448	single-end
SRR5956449	iPSC	E3	34	iPSC03APOE3	SRR5956449	single-end
SRR5956450	iPSC	E4	37	iPSC04APOE4	SRR5956450	single-end
SRR5956451	iPSC	E4	38	iPSC05APOE4	SRR5956451	single-end
SRR5956452	iPSC	E4	39	iPSC06APOE4	SRR5956452	single-end
SRR5956453	astrocyte	E3	4	AST07APOE3	SRR5956453	single-end
SRR5956454	astrocyte	E3	4	AST08APOE3	SRR5956454	single-end
SRR5956455	astrocyte	E3	4	AST09APOE3	SRR5956455	single-end
SRR5956456	astrocyte	E4	4	AST10APOE4	SRR5956456	single-end
SRR5956457	astrocyte	E4	4	AST11APOE4	SRR5956457	single-end
SRR5956458	astrocyte	E4	4	AST12APOE4	SRR5956458	single-end
SRR5956459	microglia_like	E3	1	MIC13APOE3	SRR5956459	paired-end
SRR5956460	microglia_like	E3	1	MIC14APOE3	SRR5956460	paired-end
SRR5956461	microglia_like	E3	1	MIC15APOE3	SRR5956461	paired-end

SRR5956462	microglia_like	E3	1	MIC16APOE3	SRR5956462	paired-end
SRR5956463	microglia_like	E4	1	MIC17APOE4	SRR5956463	paired-end
SRR5956464	microglia_like	E4	1	MIC18APOE4	SRR5956464	paired-end
SRR5956465	microglia_like	E4	1	MIC19APOE4	SRR5956465	paired-end
SRR5956466	microglia_like	E4	1	MIC20APOE4	SRR5956466	paired-end
SRR5956467	neuron	E3	1	NEU21APOE3	SRR5956467	single-end
SRR5956468	neuron	E3	1	NEU22APOE3	SRR5956468	single-end
SRR5956469	neuron	E3	1	NEU23APOE3	SRR5956469	single-end
SRR5956470	neuron	E4	1	NEU24APOE4	SRR5956470	single-end
SRR5956471	neuron	E4	1	NEU25APOE4	SRR5956471	single-end
SRR5956472	neuron	E4	1	NEU26APOE4	SRR5956472	single-end

7.1.4. PRJNA451437

Meta information: Illumina Hi-seq 2500, olfactory bulb

Overall design: RNA-Seq of the olfactory bulb of 19 AD patients and 20 age-matched controls were generated by deep sequencing postmortem tissue.

Table 12 - Design matrix of PRJNA451437

	tissue	condition	run	type	sample
SRR7056885	postmortem olfactory bulb	healthy	SRR7056885	single-end	H1
SRR7056886	postmortem olfactory bulb	healthy	SRR7056886	single-end	H2
SRR7056887	postmortem olfactory bulb	healthy	SRR7056887	single-end	H3
SRR7056888	postmortem olfactory bulb	healthy	SRR7056888	single-end	H4
SRR7056889	postmortem olfactory bulb	healthy	SRR7056889	single-end	H5
SRR7056890	postmortem olfactory bulb	healthy	SRR7056890	single-end	H6
SRR7056891	postmortem olfactory bulb	healthy	SRR7056891	single-end	H7
SRR7056892	postmortem olfactory bulb	healthy	SRR7056892	single-end	H8
SRR7056893	postmortem olfactory bulb	healthy	SRR7056893	single-end	H9
SRR7056894	postmortem olfactory bulb	healthy	SRR7056894	single-end	H10
SRR7056895	postmortem olfactory bulb	AD	SRR7056895	single-end	A1
SRR7056896	postmortem olfactory bulb	AD	SRR7056896	single-end	A2
SRR7056897	postmortem olfactory bulb	AD	SRR7056897	single-end	A3
SRR7056898	postmortem olfactory bulb	AD	SRR7056898	single-end	A4
SRR7056899	postmortem olfactory bulb	AD	SRR7056899	single-end	A5
SRR7056900	postmortem olfactory bulb	AD	SRR7056900	single-end	A6
SRR7056901	postmortem olfactory bulb	AD	SRR7056901	single-end	A7
SRR7056902	postmortem olfactory bulb	AD	SRR7056902	single-end	A8
SRR7056903	postmortem olfactory bulb	AD	SRR7056903	single-end	A9
SRR7056904	postmortem olfactory bulb	AD	SRR7056904	single-end	A10
SRR8942860	postmortem olfactory bulb	AD	SRR8942860	single-end	A11
SRR8942861	postmortem olfactory bulb	healthy	SRR8942861	single-end	H11
SRR8942862	postmortem olfactory bulb	healthy	SRR8942862	single-end	H12
SRR8942863	postmortem olfactory bulb	AD	SRR8942863	single-end	A12
SRR8942864	postmortem olfactory bulb	healthy	SRR8942864	single-end	H13
SRR8942865	postmortem olfactory bulb	healthy	SRR8942865	single-end	H14
SRR8942866	postmortem olfactory bulb	healthy	SRR8942866	single-end	H15
SRR8942867	postmortem olfactory bulb	AD	SRR8942867	single-end	A13

SRR8942868	postmortem olfactory bulb	healthy	SRR8942868	single-end	H16
SRR8942869	postmortem olfactory bulb	AD	SRR8942869	single-end	A14
SRR8942870	postmortem olfactory bulb	AD	SRR8942870	single-end	A15
SRR8942871	postmortem olfactory bulb	AD	SRR8942871	single-end	A16
SRR8942872	postmortem olfactory bulb	AD	SRR8942872	single-end	A17
SRR8942873	postmortem olfactory bulb	AD	SRR8942873	single-end	A18
SRR8942874	postmortem olfactory bulb	healthy	SRR8942874	single-end	H17
SRR8942875	postmortem olfactory bulb	AD	SRR8942875	single-end	A19
SRR8942876	postmortem olfactory bulb	healthy	SRR8942876	single-end	H18
SRR8942877	postmortem olfactory bulb	healthy	SRR8942877	single-end	H19
SRR8942878	postmortem olfactory bulb	healthy	SRR8942878	single-end	H20

7.1.5. PRJNA482601

Meta information: Illumina HiSeq 2500, iPSC-derived cerebral organoid

Overall design: Exploring the effects of isogenic editing of APOE E4 to E3 in cerebral organoids. Comparison of APOE E4 vs. E3 isogenic organoids with 3 biological replicates per group.

Table 13 - Design matrix of PRJNA482601

	condition	run	type	sample	cell
SRR7586681	E4E4	SRR7586681	paired-end	AD1	cerebral
SRR7586682	E4E4	SRR7586682	paired-end	AD2	cerebral
SRR7586683	E3E3	SRR7586683	paired-end	H1	cerebral
SRR7586684	E3E3	SRR7586684	paired-end	H2	cerebral
SRR7586685	E4E4	SRR7586685	paired-end	AD3	cerebral
SRR7586686	E3E3	SRR7586686	paired-end	H3	cerebral

7.1.6. PRJNA527202

Meta information: Illumina HiSeq 2500, iPSC-derived neurons

Overall design: RNAseq analysis of wildtype and three different mutated human iPS cells differentiated into neurons

Table 14 - Design matrix of PRJNA527202

	cell	condition	run	type	sample
SRR8732095	iPSC-derived_neuron	WT	SRR8732095	paired-end	WT1
SRR8732096	iPSC-derived_neuron	WT	SRR8732096	paired-end	WT2
SRR8732097	iPSC-derived_neuron	WT	SRR8732097	paired-end	WT3
SRR8732098	iPSC-derived_neuron	APP	SRR8732098	paired-end	A1
SRR8732099	iPSC-derived_neuron	APP	SRR8732099	paired-end	A2
SRR8732100	iPSC-derived_neuron	APP	SRR8732100	paired-end	A3
SRR8732101	iPSC-derived_neuron	PSEN1	SRR8732101	paired-end	P1
SRR8732102	iPSC-derived_neuron	PSEN1	SRR8732102	paired-end	P2
SRR8732103	iPSC-derived_neuron	PSEN1	SRR8732103	paired-end	P3
SRR8732104	iPSC-derived_neuron	APP_PSEN1	SRR8732104	paired-end	AP1

SRR8732105	iPSC-derived_neuron	APP_PSEN1	SRR8732105	paired-end	AP2
SRR8732106	iPSC-derived_neuron	APP_PSEN1	SRR8732106	paired-end	AP3

7.1.7. PRJNA559812

Meta information: Illumina HiSeq 3000, iPSC-derived microglia

Overall design: Human iPSC-derived microglia-like cells were collected and analyzed for gene expression using RNA-seq

Table 15 - Design matrix of PRJNA559812

	variant	cell	sample	run	type
SRR9951048	PSEN1_deltaE9	IPSC-microglia	PD1	SRR9951048	single-end
SRR9951049	PSEN1_deltaE9_corrected	IPSC-microglia	IPDC2	SRR9951049	single-end
SRR9951050	PSEN1_deltaE9	IPSC-microglia	PD3	SRR9951050	single-end
SRR9951051	PSEN1_deltaE9_corrected	IPSC-microglia	IPDC4	SRR9951051	single-end
SRR9951052	PSEN1_deltaE9	IPSC-microglia	PD1	SRR9951052	single-end
SRR9951053	PSEN1_deltaE9_corrected	IPSC-microglia	IPDC2	SRR9951053	single-end
SRR9951054	PSEN1_deltaE9	IPSC-microglia	PD3	SRR9951054	single-end
SRR9951055	PSEN1_deltaE9_corrected	IPSC-microglia	IPDC4	SRR9951055	single-end
SRR9951056	PSEN1_deltaE9	IPSC-microglia	PD1	SRR9951056	single-end
SRR9951057	PSEN1_deltaE9_corrected	IPSC-microglia	IPDC2	SRR9951057	single-end
SRR9951058	PSEN1_deltaE9	IPSC-microglia	PD3	SRR9951058	single-end
SRR9951059	PSEN1_deltaE9_corrected	IPSC-microglia	IPDC4	SRR9951059	single-end

7.1.8. PRJNA576835

Meta information: Illumina NextSeq 500, iPSC-derived astrocytes

Overall design: Ribo-depleted RNA was sequenced with Illumina Next-seq 500 from iPS-derived astrocytes

Table 16 - Design matrix of PRJNA576835

	run	type	PSEN1	sample	cell	sex	APOE	condition
SRR10257345	SRR10257345	single-end	delta_E9	AD1M	hiPSC-derived-brain-astrocytes	male	e3/e3	AD
SRR10257346	SRR10257346	single-end	delta_E9	AD1M	hiPSC-derived-brain-astrocytes	male	e3/e3	AD
SRR10257347	SRR10257347	single-end	delta_E9	AD1M	hiPSC-derived-brain-astrocytes	male	e3/e3	AD
SRR10257348	SRR10257348	single-end	corrected	MC2M	hiPSC-derived-brain-astrocytes	male	e3/e3	mutation-corrected
SRR10257349	SRR10257349	single-end	corrected	MC2M	hiPSC-derived-brain-astrocytes	male	e3/e3	mutation-corrected
SRR10257350	SRR10257350	single-end	corrected	MC2M	hiPSC-derived-brain-astrocytes	male	e3/e3	mutation-corrected
SRR10257351	SRR10257351	single-end	delta_E9	PS3F	hiPSC-derived-brain-astrocytes	female	e3/e3	AD
SRR10257352	SRR10257352	single-end	delta_E9	PS3F	hiPSC-derived-brain-astrocytes	female	e3/e3	AD
SRR10257353	SRR10257353	single-	delta_E9	PS3F	hiPSC-derived-brain-	female	e3/e3	AD

		end			astrocytes			
SRR10257354	SRR10257354	single-end	corrected	MC4F	hiPSC-derived-brain-astrocytes	female	e3/e3	mutation-corrected
SRR10257355	SRR10257355	single-end	corrected	MC4F	hiPSC-derived-brain-astrocytes	female	e3/e3	mutation-corrected
SRR10257356	SRR10257356	single-end	corrected	MC4F	hiPSC-derived-brain-astrocytes	female	e3/e3	mutation-corrected

7.1.9. PRJNA603192

Meta information: Illumina HiSeq 2000, inferior paritel lobule

Overall design: Postmortem brain samples from inferior parietal lobule genotyped as APOEε2/c (APOEε2/carriers), APOEε3/3, and APOEε4/c (APOEε4/carriers), age- and gender-matched, were used to reveal APOE allele-associated changes in transcriptomes and lipidomes

Table 17 - Design matrix of PRJNA603192

	run	type	tissue	sex	APOE	sample
SRR10964980	SRR10964980	single-end	inferior-parietal-lobule	female	E2c	1-2F
SRR10964981	SRR10964981	single-end	inferior-parietal-lobule	female	E2c	2-2F
SRR10964982	SRR10964982	single-end	inferior-parietal-lobule	female	E33	3-3F
SRR10964983	SRR10964983	single-end	inferior-parietal-lobule	male	E2c	4-2M
SRR10964984	SRR10964984	single-end	inferior-parietal-lobule	female	E4c	5-4F
SRR10964985	SRR10964985	single-end	inferior-parietal-lobule	female	E4c	6-4F
SRR10964986	SRR10964986	single-end	inferior-parietal-lobule	male	E4c	7-4M
SRR10964987	SRR10964987	single-end	inferior-parietal-lobule	female	E2c	8-2F
SRR10964988	SRR10964988	single-end	inferior-parietal-lobule	female	E4c	9-4F
SRR10964989	SRR10964989	single-end	inferior-parietal-lobule	male	E4c	10-4M
SRR10964990	SRR10964990	single-end	inferior-parietal-lobule	male	E4c	11-4M
SRR10964991	SRR10964991	single-end	inferior-parietal-lobule	male	E4c	12-4M
SRR10964992	SRR10964992	single-end	inferior-parietal-lobule	female	E33	13-3F
SRR10964993	SRR10964993	single-end	inferior-parietal-lobule	male	E2c	14-2M
SRR10964994	SRR10964994	single-end	inferior-parietal-lobule	female	E2c	15-2F
SRR10964995	SRR10964995	single-end	inferior-parietal-lobule	female	E2c	16-2F
SRR10964996	SRR10964996	single-end	inferior-parietal-lobule	male	E2c	17-2M
SRR10964997	SRR10964997	single-end	inferior-parietal-lobule	female	E33	18-3F
SRR10964998	SRR10964998	single-end	inferior-parietal-lobule	male	E33	19-3M
SRR10964999	SRR10964999	single-end	inferior-parietal-lobule	male	E33	20-3M
SRR10965000	SRR10965000	single-end	inferior-parietal-lobule	female	E4c	21-4F
SRR10965001	SRR10965001	single-end	inferior-parietal-lobule	female	E33	22-3F
SRR10965002	SRR10965002	single-end	inferior-parietal-lobule	male	E4c	23-4M
SRR10965003	SRR10965003	single-end	inferior-parietal-lobule	male	E33	24-3M
SRR10965004	SRR10965004	single-end	inferior-parietal-lobule	male	E4c	25-4M
SRR10965005	SRR10965005	single-end	inferior-parietal-lobule	male	E33	26-3M
SRR10965006	SRR10965006	single-end	inferior-parietal-lobule	male	E4c	27-4M
SRR10965007	SRR10965007	single-end	inferior-parietal-lobule	female	E4c	28-4M
SRR10965008	SRR10965008	single-end	inferior-parietal-lobule	female	E33	29-3F
SRR10965009	SRR10965009	single-end	inferior-parietal-lobule	male	E33	30-3M

SRR10965010	SRR10965010	single-end	inferior-parietal-lobule	male	E4c	31-4M
SRR10965011	SRR10965011	single-end	inferior-parietal-lobule	male	E4c	32-4M
SRR10965012	SRR10965012	single-end	inferior-parietal-lobule	female	E4c	33-4F
SRR10965013	SRR10965013	single-end	inferior-parietal-lobule	male	E4c	34-4M
SRR10965014	SRR10965014	single-end	inferior-parietal-lobule	male	E4c	35-4M
SRR10965015	SRR10965015	single-end	inferior-parietal-lobule	female	E4c	36-4F
SRR10965016	SRR10965016	single-end	inferior-parietal-lobule	male	E4c	37-4M
SRR10965017	SRR10965017	single-end	inferior-parietal-lobule	female	E4c	38-4F
SRR10965018	SRR10965018	single-end	inferior-parietal-lobule	female	E4c	39-4F
SRR10965019	SRR10965019	single-end	inferior-parietal-lobule	female	E33	40-3F
SRR10965020	SRR10965020	single-end	inferior-parietal-lobule	female	E4c	41-4F
SRR10965021	SRR10965021	single-end	inferior-parietal-lobule	female	E33	42-3F

7.1.10. PRJNA643561

Meta information: NextSeq 500, hESC-derived microglia

Overall design: Examination of gene expression profiles of isogenic human ES-derived microglia-like cell lines (hMGLs) harboring AD variants in CD33, INPP5D, SORL1, and TREM2 loci

Table 18 - Design matrix of PRJNA643561

	cell	condition	sample	run	type
SRR12129906	hMGLs	WT	WT	SRR12129906	paired-end
SRR12129907	hMGLs	TREM2_KO	TREM2-KO	SRR12129907	paired-end
SRR12129908	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129908	paired-end
SRR12129909	hMGLs	SORL1_KO	SORL1-KO	SRR12129909	paired-end
SRR12129910	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129910	paired-end
SRR12129911	hMGLs	WT	WT	SRR12129911	paired-end
SRR12129912	hMGLs	TREM2_KO	TREM2-KO	SRR12129912	paired-end
SRR12129913	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129913	paired-end
SRR12129914	hMGLs	SORL1_KO	SORL1-KO	SRR12129914	paired-end
SRR12129915	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129915	paired-end
SRR12129916	hMGLs	WT	WT	SRR12129916	paired-end
SRR12129917	hMGLs	TREM2_KO	TREM2-KO	SRR12129917	paired-end
SRR12129918	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129918	paired-end
SRR12129919	hMGLs	SORL1_KO	SORL1-KO	SRR12129919	paired-end
SRR12129920	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129920	paired-end
SRR12129921	hMGLs	WT	WT	SRR12129921	paired-end
SRR12129922	hMGLs	CD33_SNP	CD33 SNP	SRR12129922	paired-end
SRR12129923	hMGLs	TREM2_KO	TREM2-KO	SRR12129923	paired-end
SRR12129924	hMGLs	INPP5D_SNP	INPP5D SNP	SRR12129924	paired-end
SRR12129925	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129925	paired-end
SRR12129926	hMGLs	SORL1_KO	SORL1-KO	SRR12129926	paired-end
SRR12129927	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129927	paired-end
SRR12129928	hMGLs	WT	WT	SRR12129928	paired-end
SRR12129929	hMGLs	CD33_SNP	CD33 SNP	SRR12129929	paired-end
SRR12129930	hMGLs	TREM2_KO	TREM2-KO	SRR12129930	paired-end
SRR12129931	hMGLs	INPP5D_SNP	INPP5D SNP	SRR12129931	paired-end

SRR12129932	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129932	paired-end
SRR12129933	hMGLs	SORL1_KO	SORL1-KO	SRR12129933	paired-end
SRR12129934	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129934	paired-end
SRR12129935	hMGLs	WT	WT	SRR12129935	paired-end
SRR12129936	hMGLs	TREM2_KO	TREM2-KO	SRR12129936	paired-end
SRR12129937	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129937	paired-end
SRR12129938	hMGLs	SORL1_KO	SORL1-KO	SRR12129938	paired-end
SRR12129939	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129939	paired-end
SRR12129940	hMGLs	WT	WT	SRR12129940	paired-end
SRR12129941	hMGLs	TREM2_KO	TREM2-KO	SRR12129941	paired-end
SRR12129942	hMGLs	TREM2_R47H	TREM2-R47H	SRR12129942	paired-end
SRR12129943	hMGLs	TREM2_KO	TREM2-KO	SRR12129943	paired-end
SRR12129944	hMGLs	SORL1_KO	SORL1-KO	SRR12129944	paired-end
SRR12129945	hMGLs	SORL1_A528T	SORL1-A528T	SRR12129945	paired-end

7.1.11. PRJNA644383

Meta information: NextSeq 500, lateral temporal lobe

Overall design: ERCC spike-in controlled RNA-seq in postmortal lateral temporal lobe of Alzheimer's disease affected brains (AD), control Old (Old) and control Young (Young)

Table 19 - Design matrix of PRJNA644383

	tissue	age	condition	run	type	sample
SRR12158131	lateral temporal lobe	old	healthy	SRR12158131	single-end	H1Old
SRR12158132	lateral temporal lobe	old	healthy	SRR12158132	single-end	H2Old
SRR12158133	lateral temporal lobe	old	AD	SRR12158133	single-end	A1Old
SRR12158134	lateral temporal lobe	old	healthy	SRR12158134	single-end	H3Old
SRR12158135	lateral temporal lobe	young	healthy	SRR12158135	single-end	H4Yng
SRR12158136	lateral temporal lobe	young	healthy	SRR12158136	single-end	H5Yng
SRR12158137	lateral temporal lobe	young	healthy	SRR12158137	single-end	H6Yng
SRR12158138	lateral temporal lobe	old	AD	SRR12158138	single-end	A2Old
SRR12158139	lateral temporal lobe	old	AD	SRR12158139	single-end	A3Old
SRR12158140	lateral temporal lobe	old	AD	SRR12158140	single-end	A4Old
SRR12158141	lateral temporal lobe	old	AD	SRR12158141	single-end	A5Old
SRR12158142	lateral temporal lobe	old	healthy	SRR12158142	single-end	H7Old
SRR12158143	lateral temporal lobe	old	healthy	SRR12158143	single-end	H8Old
SRR12158144	lateral temporal lobe	old	healthy	SRR12158144	single-end	H9Old
SRR12158145	lateral temporal lobe	old	AD	SRR12158145	single-end	A6Old
SRR12158146	lateral temporal lobe	old	healthy	SRR12158146	single-end	H10Old
SRR12158147	lateral temporal lobe	old	healthy	SRR12158147	single-end	H11Old
SRR12158148	lateral temporal lobe	old	healthy	SRR12158148	single-end	H12Old
SRR12158149	lateral temporal lobe	young	healthy	SRR12158149	single-end	H13Yng
SRR12158150	lateral temporal lobe	young	healthy	SRR12158150	single-end	H14Yng
SRR12158151	lateral temporal lobe	young	healthy	SRR12158151	single-end	H15Yng
SRR12158152	lateral temporal lobe	young	healthy	SRR12158152	single-end	H16Yng
SRR12158153	lateral temporal lobe	young	healthy	SRR12158153	single-end	H17Yng
SRR12158154	lateral temporal lobe	old	AD	SRR12158154	single-end	A7Old

SRR12158155	lateral temporal lobe	old	AD	SRR12158155	single-end	A8Old
SRR12158156	lateral temporal lobe	old	AD	SRR12158156	single-end	A9Old
SRR12158157	lateral temporal lobe	old	AD	SRR12158157	single-end	A10Old
SRR12158158	lateral temporal lobe	old	AD	SRR12158158	single-end	A11Old
SRR12158159	lateral temporal lobe	old	AD	SRR12158159	single-end	A12Old
SRR12158160	lateral temporal lobe	old	healthy	SRR12158160	single-end	H18Old

7.1.12. PRJNA662330

Meta information: Illumina HiSeq 4000, iPSC-derived microglia

Overall design: Comparison of WT and KO TREM2 microglia at baseline and after treatment

Table 20 - Design matrix of PRJNA662330

	cell	condition	treatment	run	type	sample
SRR12608405	microglia from iPSCs	WT	untreated	SRR12608405	paired-end	1WU
SRR12608406	microglia from iPSCs	WT	untreated	SRR12608406	paired-end	2WU
SRR12608407	microglia from iPSCs	WT	untreated	SRR12608407	paired-end	3WU
SRR12608408	microglia from iPSCs	WT	untreated	SRR12608408	paired-end	4WU
SRR12608409	microglia from iPSCs	KO	untreated	SRR12608409	paired-end	5KU
SRR12608410	microglia from iPSCs	KO	untreated	SRR12608410	paired-end	6KU
SRR12608411	microglia from iPSCs	KO	untreated	SRR12608411	paired-end	7KU
SRR12608412	microglia from iPSCs	KO	untreated	SRR12608412	paired-end	8KU

7.1.13. PRJNA675864

Meta information: NextSeq 500, blood

Overall design: Peripheral Blood Mononuclear Cells derived from 6 SALS, 6 AD, and 6 PD patients and 14 age- and sex-matched healthy controls (CTR) were deep-sequenced and included in Real-Time PCR experiments.

Table 21 - Design matrix of PRJNA675864

	tissue	condition	cell	type	sample
SRR13019394	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H1
SRR13019395	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H2
SRR13019396	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H3
SRR13019397	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H4
SRR13019404	blood	AD	peripheral-blood-mononuclear-cell	paired-end	A1
SRR13019405	blood	AD	peripheral-blood-mononuclear-cell	paired-end	A2
SRR13019406	blood	AD	peripheral-blood-mononuclear-cell	paired-end	A3
SRR13019407	blood	AD	peripheral-blood-mononuclear-cell	paired-end	A4
SRR13019408	blood	AD	peripheral-blood-mononuclear-cell	paired-end	A5
SRR13019409	blood	AD	peripheral-blood-mononuclear-cell	paired-end	A6
SRR13019410	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H5
SRR13019411	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H6
SRR13019412	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H7

SRR13019413	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H8
SRR13019414	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H9
SRR13019415	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H10
SRR13019416	blood	healthy	peripheral-blood-mononuclear-cell	paired-end	H11

7.1.14. PRJNA683625

Meta information: Illumina HiSeq 2000, iPSC-derived neural progenitor cells

Overall design: Alzheimer's disease vs. healthy controls

Table 22 - Design matrix of PRJNA683625

	condition	sample	cells	run	type
SRR13219955	AD	AD1	neural progenitor cells	SRR13219955	paired-end
SRR13219956	AD	AD2	neural progenitor cells	SRR13219956	paired-end
SRR13219957	AD	AD3	neural progenitor cells	SRR13219957	paired-end
SRR13219958	AD	AD4	neural progenitor cells	SRR13219958	paired-end
SRR13219959	healthy	H1	neural progenitor cells	SRR13219959	paired-end
SRR13219960	healthy	H2	neural progenitor cells	SRR13219960	paired-end
SRR13219961	healthy	H3	neural progenitor cells	SRR13219961	paired-end
SRR13219962	healthy	H4	neural progenitor cells	SRR13219962	paired-end

7.1.15. PRJNA688060

Meta information: Ion Torrent Proton, Middle temporal gyrus

Overall design: Total RNA was extracted from human post-mortem brain samples (Middle temporal gyrus; 3 AD, 4 healthy control), and a modified RNA-Seq protocol used to simultaneously examine coding and non-coding RNA

Table 23 - Design matrix of PRJNA688060

	run	type	tissue	condition	sample
SRR13310019	SRR13310019	single-end	middle-temporal-gyrus	AD	A1
SRR13310020	SRR13310020	single-end	middle-temporal-gyrus	healthy	H1
SRR13310021	SRR13310021	single-end	middle-temporal-gyrus	healthy	H2
SRR13310022	SRR13310022	single-end	middle-temporal-gyrus	healthy	H3
SRR13310023	SRR13310023	single-end	middle-temporal-gyrus	AD	A2
SRR13310024	SRR13310024	single-end	middle-temporal-gyrus	healthy	H4
SRR13310025	SRR13310025	single-end	middle-temporal-gyrus	AD	A3

7.1.16. PRJNA688885

Meta information: Illumina NovaSeq 6000, iPSC-derived brain organoids

Overall design: 14048 brain organoids (BO2) and ADRC28 brain organoids (BO3) were treated without (control) or with 10% human serum (8748 or 3821) for 13-14 days and then were harvested at day 94-95

Table 24 - Design matrix of PRJNA688885

	run	cell	treatment	source	sample	type
SRR13336736	SRR13336736	iPSC-derived brain organoids	untreated	BO2	BO2U_1	single-end
SRR13336737	SRR13336737	iPSC-derived brain organoids	untreated	BO2	BO2U_1	single-end
SRR13336738	SRR13336738	iPSC-derived brain organoids	untreated	BO2	BO2U_1	single-end
SRR13336739	SRR13336739	iPSC-derived brain organoids	untreated	BO2	BO2U_1	single-end
SRR13336740	SRR13336740	iPSC-derived brain organoids	treated	BO2	BO2S_2	single-end
SRR13336741	SRR13336741	iPSC-derived brain organoids	treated	BO2	BO2S_2	single-end
SRR13336742	SRR13336742	iPSC-derived brain organoids	treated	BO2	BO2S_2	single-end
SRR13336743	SRR13336743	iPSC-derived brain organoids	treated	BO2	BO2S_2	single-end
SRR13336744	SRR13336744	iPSC-derived brain organoids	untreated	BO3	BO3U_3	single-end
SRR13336745	SRR13336745	iPSC-derived brain organoids	untreated	BO3	BO3U_3	single-end
SRR13336746	SRR13336746	iPSC-derived brain organoids	untreated	BO3	BO3U_3	single-end
SRR13336747	SRR13336747	iPSC-derived brain organoids	untreated	BO3	BO3U_3	single-end
SRR13336748	SRR13336748	iPSC-derived brain organoids	treated	BO3	BO3S_4	single-end
SRR13336749	SRR13336749	iPSC-derived brain organoids	treated	BO3	BO3S_4	single-end
SRR13336750	SRR13336750	iPSC-derived brain organoids	treated	BO3	BO3S_4	single-end
SRR13336751	SRR13336751	iPSC-derived brain organoids	treated	BO3	BO3S_4	single-end

7.1.17. PRJNA714081

Meta information: Ion Torrent Proton, blood

Overall design: RNA was isolated from whole blood from 5 LOAD cases and 10 controls (all female. Mean age 76.3±3.5)

Table 25 - Design matrix of PRJNA714081

	tissue	sex	condition	type	run	sample
SRR13953391	blood	female	healthy	single-end	SRR13953391	H1
SRR13953392	blood	female	healthy	single-end	SRR13953392	H2
SRR13953393	blood	female	healthy	single-end	SRR13953393	H3
SRR13953394	blood	female	healthy	single-end	SRR13953394	H4
SRR13953395	blood	female	healthy	single-end	SRR13953395	H5
SRR13953396	blood	female	healthy	single-end	SRR13953396	H6
SRR13953397	blood	female	healthy	single-end	SRR13953397	H7
SRR13953398	blood	female	healthy	single-end	SRR13953398	H8
SRR13953399	blood	female	healthy	single-end	SRR13953399	H9
SRR13953400	blood	female	healthy	single-end	SRR13953400	H10
SRR13953401	blood	female	AD	single-end	SRR13953401	A1
SRR13953402	blood	female	AD	single-end	SRR13953402	A2
SRR13953403	blood	female	AD	single-end	SRR13953403	A3
SRR13953404	blood	female	AD	single-end	SRR13953404	A4
SRR13953405	blood	female	AD	single-end	SRR13953405	A5

7.1.18. PRJNA727602

Meta information: Illumina HiSeq 1500, hippocampus

Overall design: Postmortem human hippocampus brains were analyzed from 8 AD and 10 non-AD subjects using Illumina TruSeq stranded mRNA LT Sample Prep kit. Sequences were obtained by using HiSeq1500. One AD and one non-AD sample applied two independent times to sequencer in order to obtain more reads

Table 26 - Design matrix of PRJNA727602

	tissue	sex	condition	sample	run	age	type
SRR14436589	hippocampus	female	AD	AD1	SRR14436589	88	paired-end
SRR14436590	hippocampus	female	AD	AD1	SRR14436590	88	paired-end
SRR14436591	hippocampus	female	AD	AD2	SRR14436591	95	paired-end
SRR14436592	hippocampus	female	AD	AD2	SRR14436592	95	paired-end
SRR14436593	hippocampus	female	AD	AD2	SRR14436593	95	paired-end
SRR14436594	hippocampus	female	AD	AD2	SRR14436594	95	paired-end
SRR14436595	hippocampus	female	AD	AD3	SRR14436595	95	paired-end
SRR14436596	hippocampus	female	AD	AD3	SRR14436596	95	paired-end
SRR14436597	hippocampus	female	AD	AD4	SRR14436597	95	paired-end
SRR14436598	hippocampus	female	AD	AD4	SRR14436598	95	paired-end
SRR14436599	hippocampus	male	AD	AD5	SRR14436599	99	paired-end
SRR14436600	hippocampus	male	AD	AD5	SRR14436600	99	paired-end
SRR14436601	hippocampus	male	AD	AD6	SRR14436601	83	paired-end
SRR14436602	hippocampus	male	AD	AD6	SRR14436602	83	paired-end
SRR14436603	hippocampus	male	AD	AD7	SRR14436603	90	paired-end
SRR14436604	hippocampus	male	AD	AD7	SRR14436604	90	paired-end
SRR14436605	hippocampus	female	AD	AD8	SRR14436605	84	paired-end
SRR14436606	hippocampus	female	AD	AD8	SRR14436606	84	paired-end
SRR14436607	hippocampus	female	healthy	H1	SRR14436607	87	paired-end
SRR14436608	hippocampus	female	healthy	H1	SRR14436608	87	paired-end
SRR14436609	hippocampus	female	healthy	H2	SRR14436609	80	paired-end
SRR14436610	hippocampus	female	healthy	H2	SRR14436610	80	paired-end
SRR14436611	hippocampus	female	healthy	H3	SRR14436611	84	paired-end
SRR14436612	hippocampus	female	healthy	H3	SRR14436612	84	paired-end
SRR14436613	hippocampus	male	healthy	H4	SRR14436613	77	paired-end
SRR14436614	hippocampus	male	healthy	H4	SRR14436614	77	paired-end
SRR14436615	hippocampus	male	healthy	H5	SRR14436615	55	paired-end
SRR14436616	hippocampus	male	healthy	H5	SRR14436616	55	paired-end
SRR14436617	hippocampus	male	healthy	H5	SRR14436617	55	paired-end
SRR14436618	hippocampus	male	healthy	H5	SRR14436618	55	paired-end
SRR14436619	hippocampus	female	healthy	H6	SRR14436619	72	paired-end
SRR14436620	hippocampus	female	healthy	H6	SRR14436620	72	paired-end
SRR14436621	hippocampus	female	healthy	H7	SRR14436621	78	paired-end
SRR14436622	hippocampus	female	healthy	H7	SRR14436622	78	paired-end
SRR14436623	hippocampus	male	healthy	H8	SRR14436623	83	paired-end
SRR14436624	hippocampus	male	healthy	H8	SRR14436624	83	paired-end
SRR14436625	hippocampus	male	healthy	H9	SRR14436625	80	paired-end
SRR14436626	hippocampus	male	healthy	H9	SRR14436626	80	paired-end
SRR14436627	hippocampus	male	healthy	H10	SRR14436627	74	paired-end
SRR14436628	hippocampus	male	healthy	H10	SRR14436628	74	paired-end

7.1.19. PRJNA767074

Meta information: Illumina HiSeq 2000, hippocampus

Overall design: mRNA profiles of the hippocampus in AD patients and healthy controls

Table 27 - Design matrix of PRJNA767074

	tissue	condition	type	run	sample
SRR16101430	hippocampus	healthy	paired-end	SRR16101430	H1
SRR16101431	hippocampus	healthy	paired-end	SRR16101431	H2
SRR16101432	hippocampus	healthy	paired-end	SRR16101432	H3
SRR16101433	hippocampus	healthy	paired-end	SRR16101433	H4
SRR16101434	hippocampus	healthy	paired-end	SRR16101434	H5
SRR16101435	hippocampus	AD	paired-end	SRR16101435	A1
SRR16101436	hippocampus	AD	paired-end	SRR16101436	A2
SRR16101437	hippocampus	AD	paired-end	SRR16101437	A3
SRR16101438	hippocampus	AD	paired-end	SRR16101438	A4
SRR16101439	hippocampus	AD	paired-end	SRR16101439	A5

7.2. Alignment and Annotation Results

The average number of reads in each project, along with percent alignment of each read, and annotation ratio can be seen in the following table.

Table 28 - Alignment and assignment results of trimmed data

Project ID	General Type	Reads (Million) (Mean)	Alignment% (Mean)	Assignment% (Mean)	Assigned Reads (Million) (Mean)
232669	dorsolateral_prefrontal_cortex	59.682	90.71%	41.00%	26.871
279526	hippocampus	177.600	75.96%	27.70%	64.650
399530	iPSC_derived_brain_cell	39.098	97.53%	73.03%	31.742
451437	olfactory_bulb	16.459	88.70%	29.06%	6.038
482601	iPSC_derived_cerebral_organoid	18.400	93.17%	47.52%	11.783
527202	iPSC_derived_neuron	41.250	93.17%	39.91%	22.183
559812	iPSC_derived_microglia	31.033	97.77%	64.75%	24.225
576835	iPSC_derived_astrocyte	47.658	98.23%	27.62%	18.667
603192	inferior_parietal_lobule	18.938	97.13%	35.19%	8.283
643561	hESC_derived_microglia	22.518	93.62%	72.66%	17.823
644383	lateral_temporal_lobe	37.777	94.47%	34.57%	16.250
662330	iPSC_derived_microglia	47.983	97.19%	77.70%	40.483
675864	blood	40.009	91.30%	17.95%	6.694
683625	neural_progenitor	21.775	98.70%	78.04%	18.150
688060	middle_temporal_gyrus	24.157	85.56%	28.39%	7.986
688885	iPSC_derived_cerebral_organoid	82.169	90.44%	41.45%	36.319
714081	blood	12.280	75.93%	61.51%	7.927
727602	hippocampus	16.765	99.25%	80.84%	14.315
767074	hippocampus	40.350	90.97%	36.74%	15.960

APPENDIX B

DEG ANALYSIS RESULTS OF THE COMPARISONS WITHIN PROJECTS

The following graphs were drawn using the following R packages: ggplot2 (Ginestet, 2011), pheatmap (*CRAN - Package Pheatmap*, n.d.), DESeq2, enchancedVolcano (*Bioconductor - EnhancedVolcano*, n.d.).

8.1. PRJNA232669

This dataset compares healthy and AD postmortem dorsolateral prefrontal cortex samples.

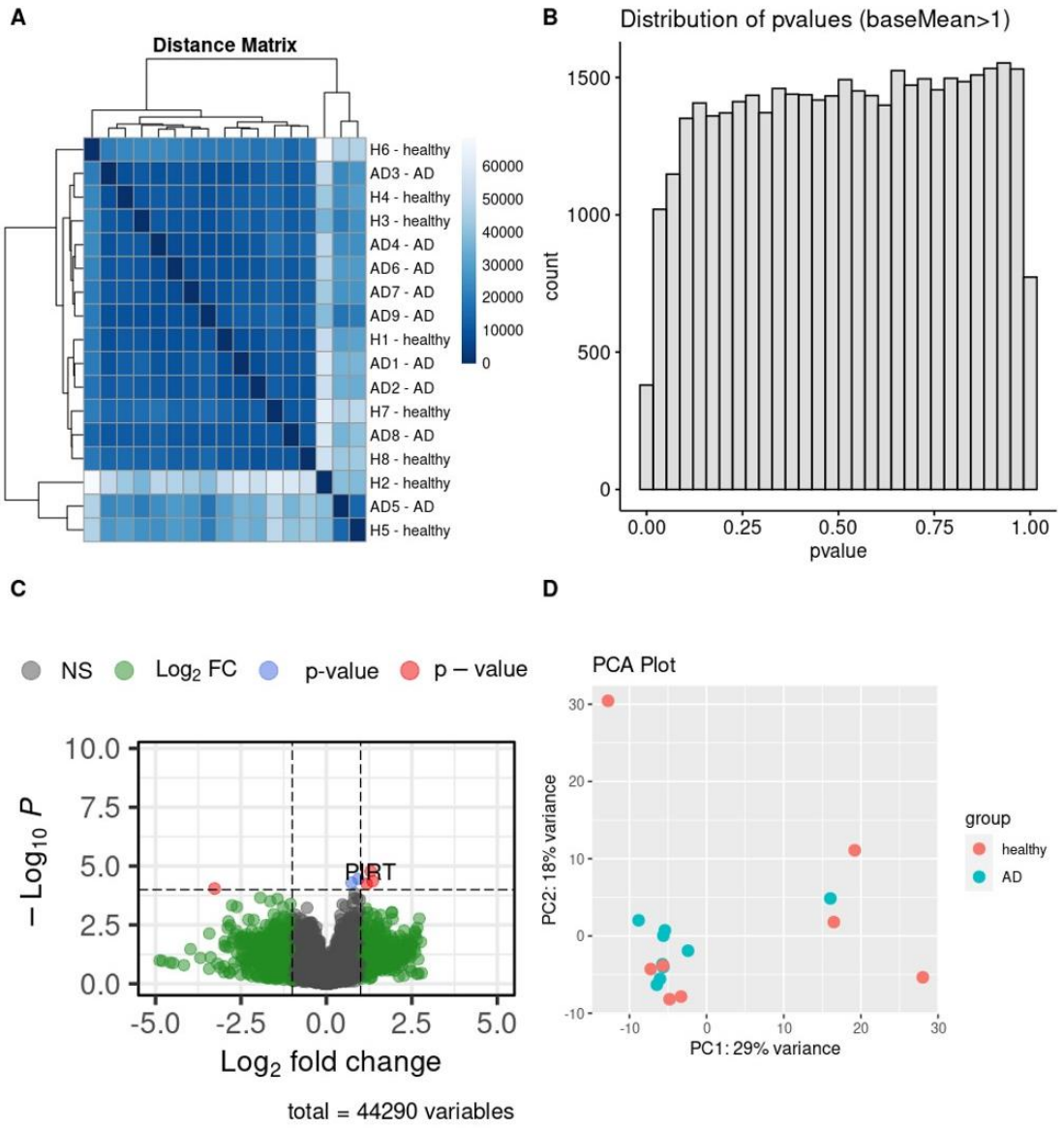

Figure 48 – Summary graphs of PRJNA232669. The significance of the gene expressions is very low and samples are not separated by the condition. As there are no significant genes (adjusted p-value < 0.05), heatmap cannot be drawn. A: Distance matrix. B: P-Value distribution C: Volcano plot D: PCA plot.

8.2. PRJNA279526

This dataset compares healthy and AD postmortem hippocampus tissue for female and male samples.

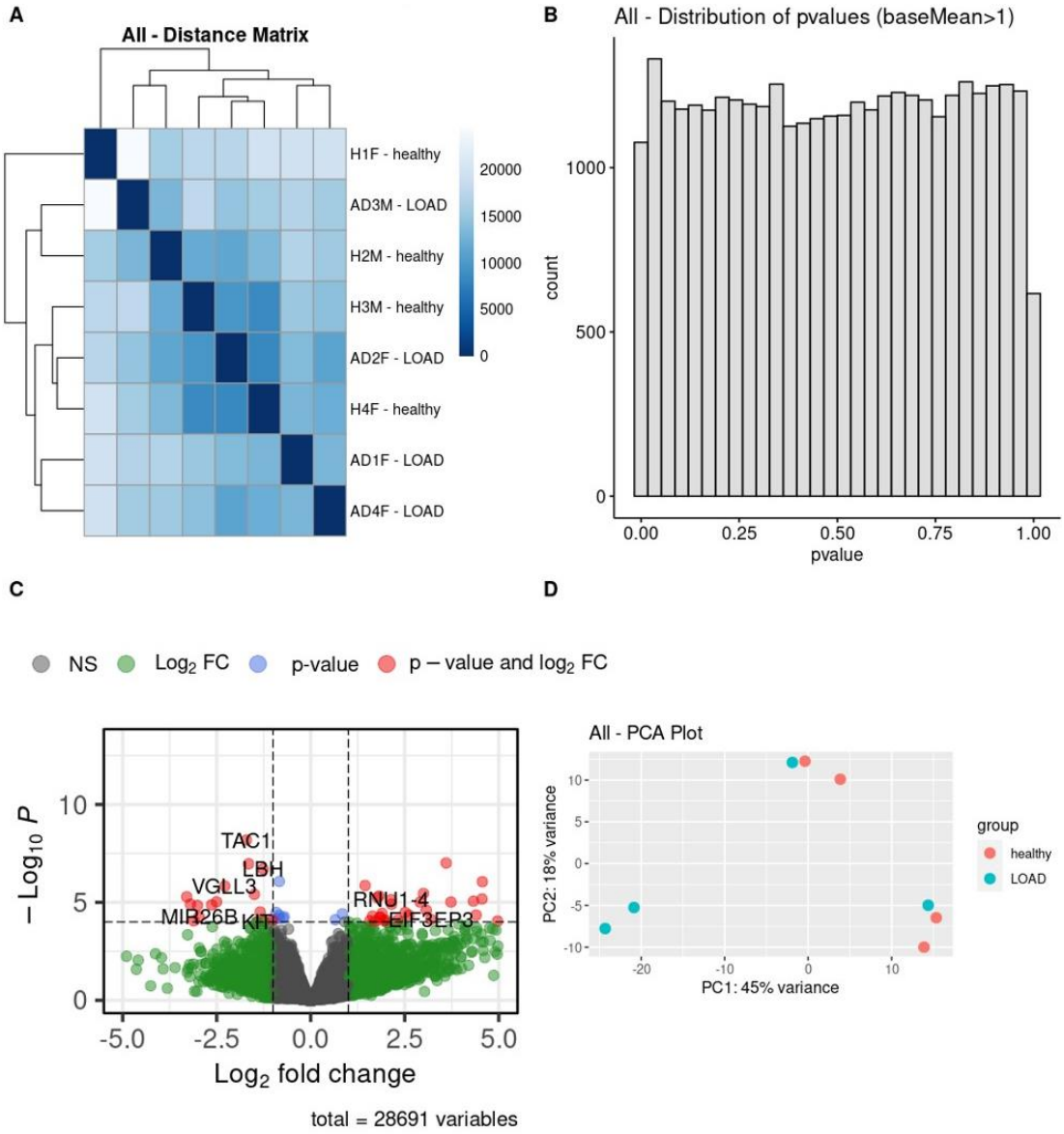

Figure 49 - Summary graphs of PRJNA279526 without sex variable. The significance of the gene expressions is low and samples are not completely separated by the condition. A: Distance matrix. B: P-Value distribution C: Volcano plot D: PCA plot.

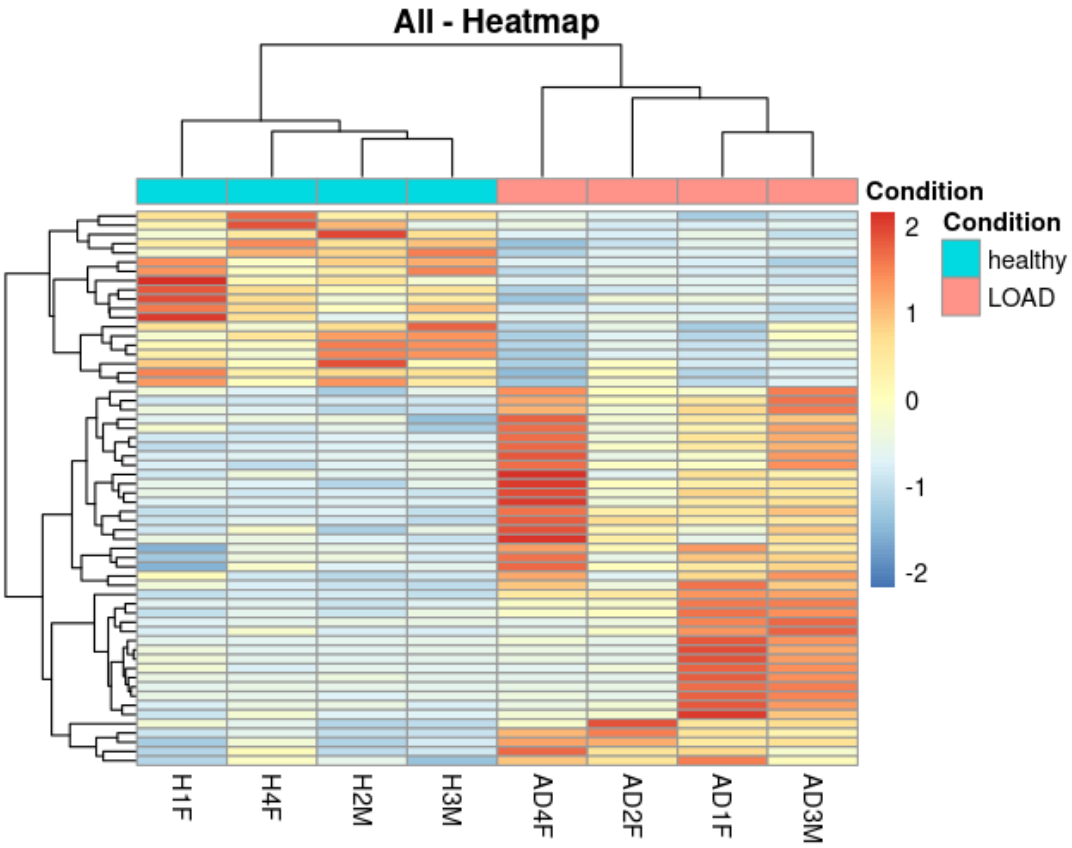

Figure 50 – Heatmap of PRJNA279526 samples. Lines represent DEGs, and the colors show expression values.

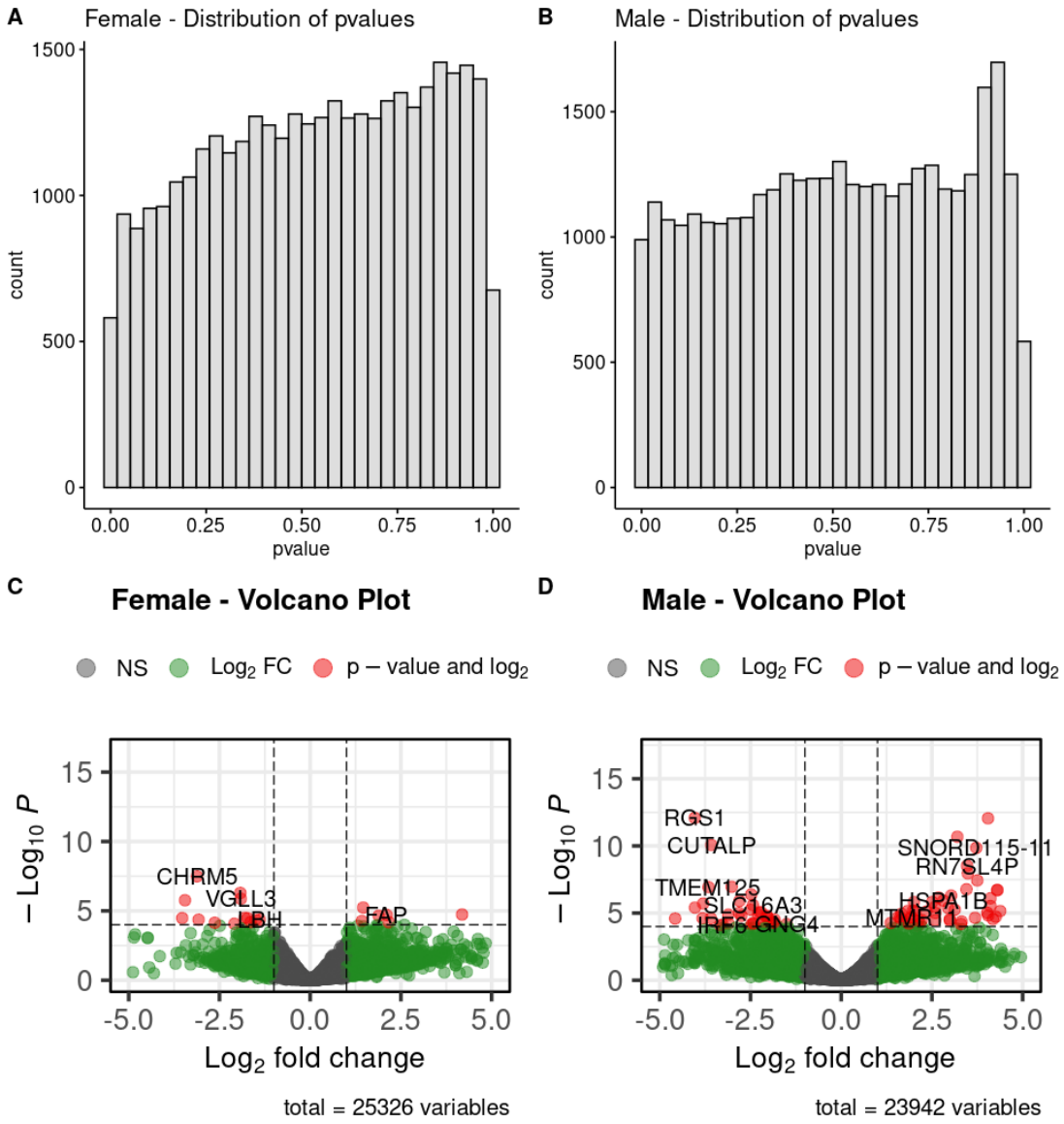

Figure 51 – P-value distributon and volcano plot of female and male samples. Male samples have a higher amount of DEGs.

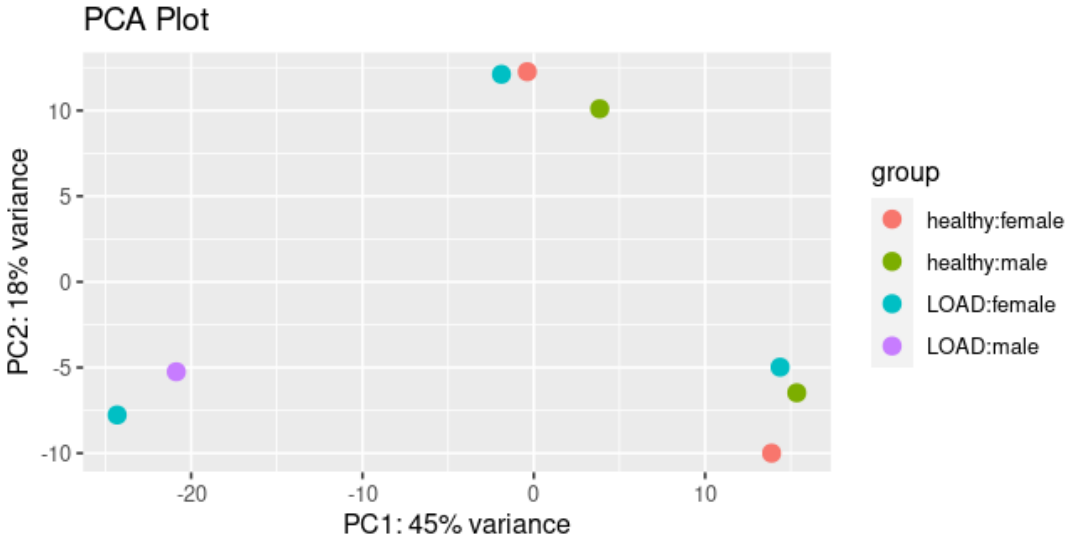

Figure 52 – PCA plot of PRJNA270529, considering sex variable. Samples are clustered but are not separated well for condition or sex.

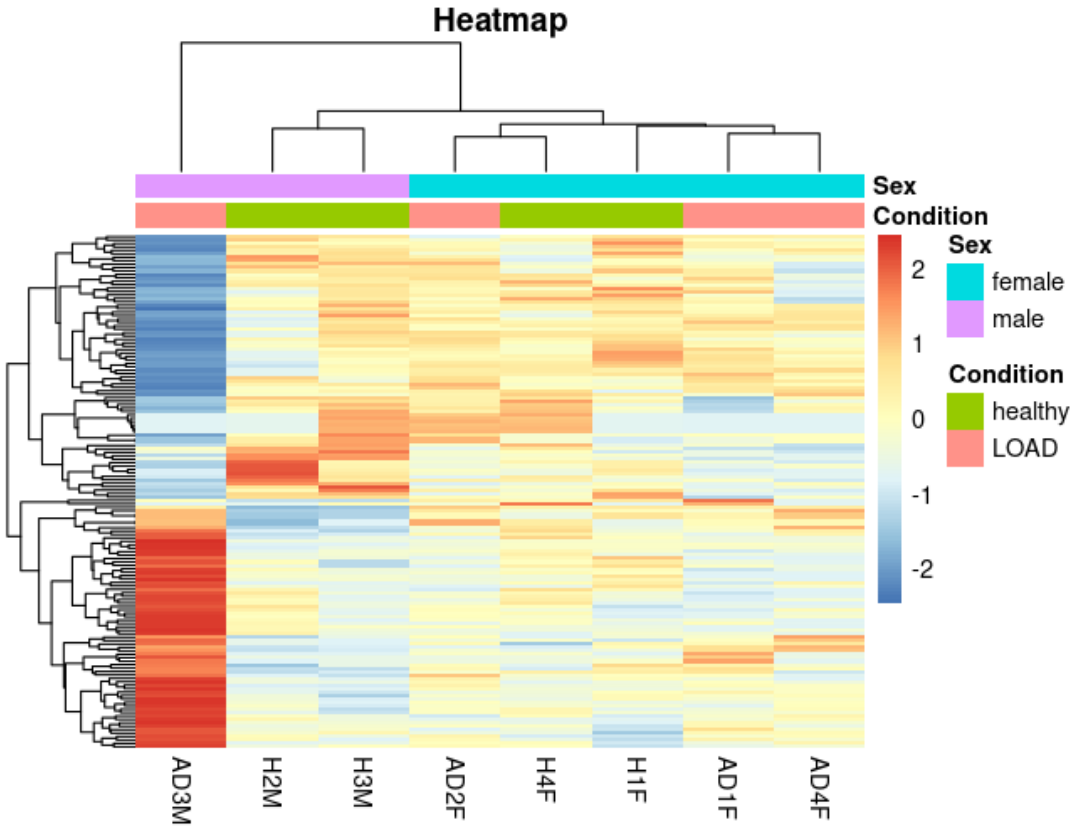

Figure 53 – Heatmap of PRJNA279526. There is no significant separation between the samples. AD3M shows abnormal expression values relative to other samples.

8.3. PRJNA399530

This dataset compares healthy (APOE3) and AD-like (APOE4) iPSC-derived brain cell types.

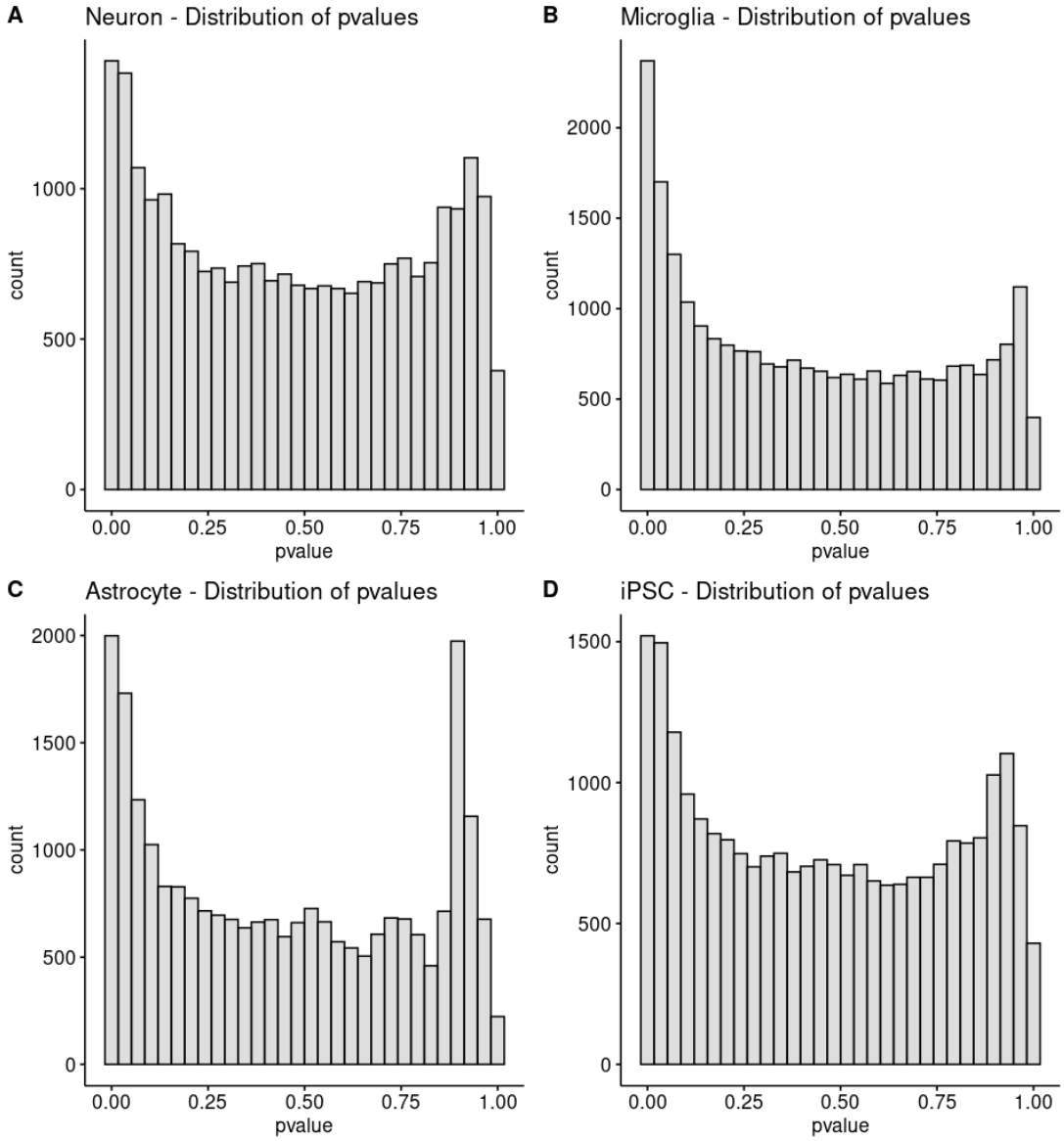

Figure 54 – P-value distribution by brain cell types. Significances are on both high and low end for all samples.

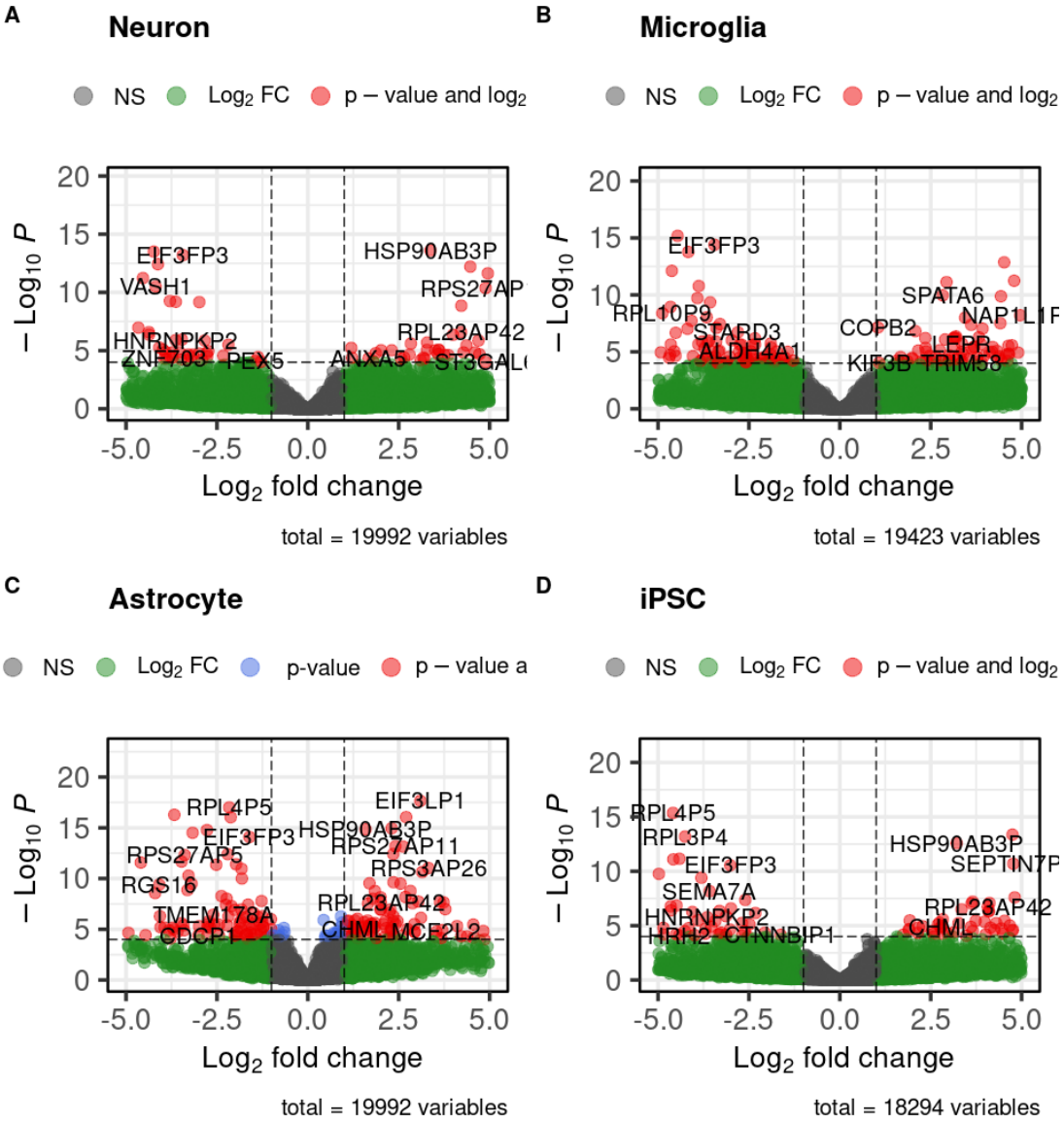

Figure 55 – Volcano plots of brain cell types. All samples have significant genes.

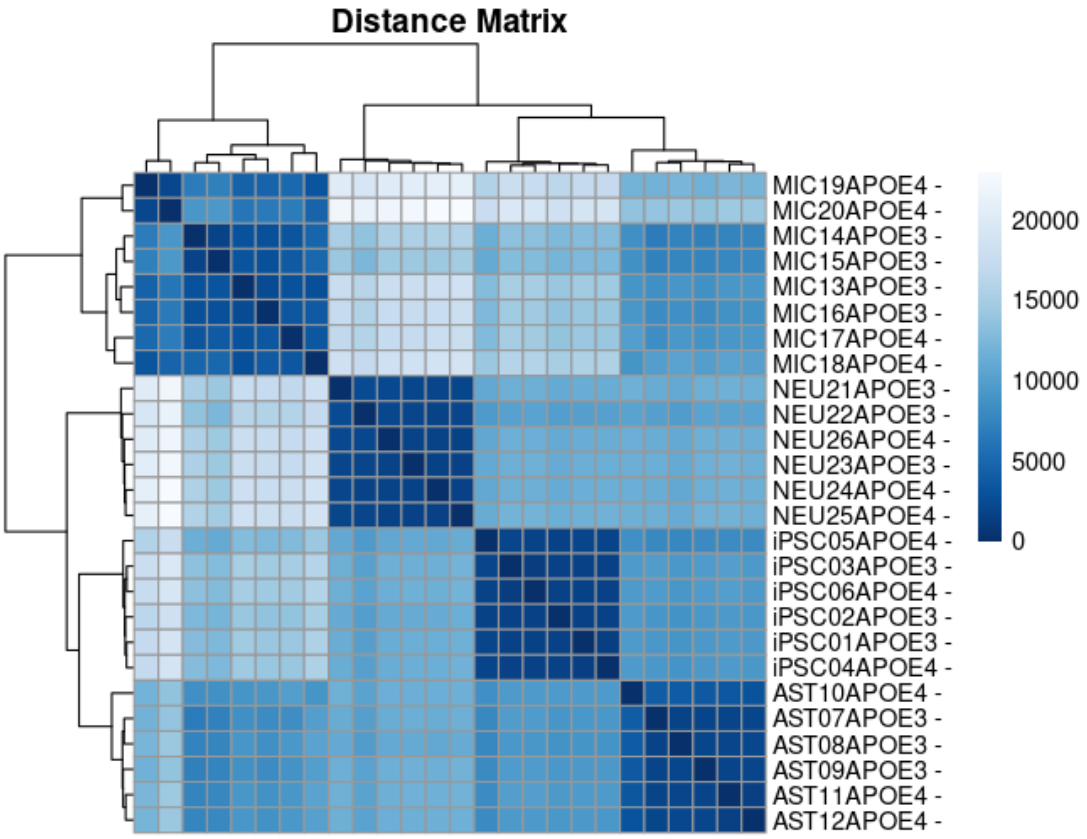

Figure 56 – Distance matrix of PRJNA399530. Samples show a visible clustering.

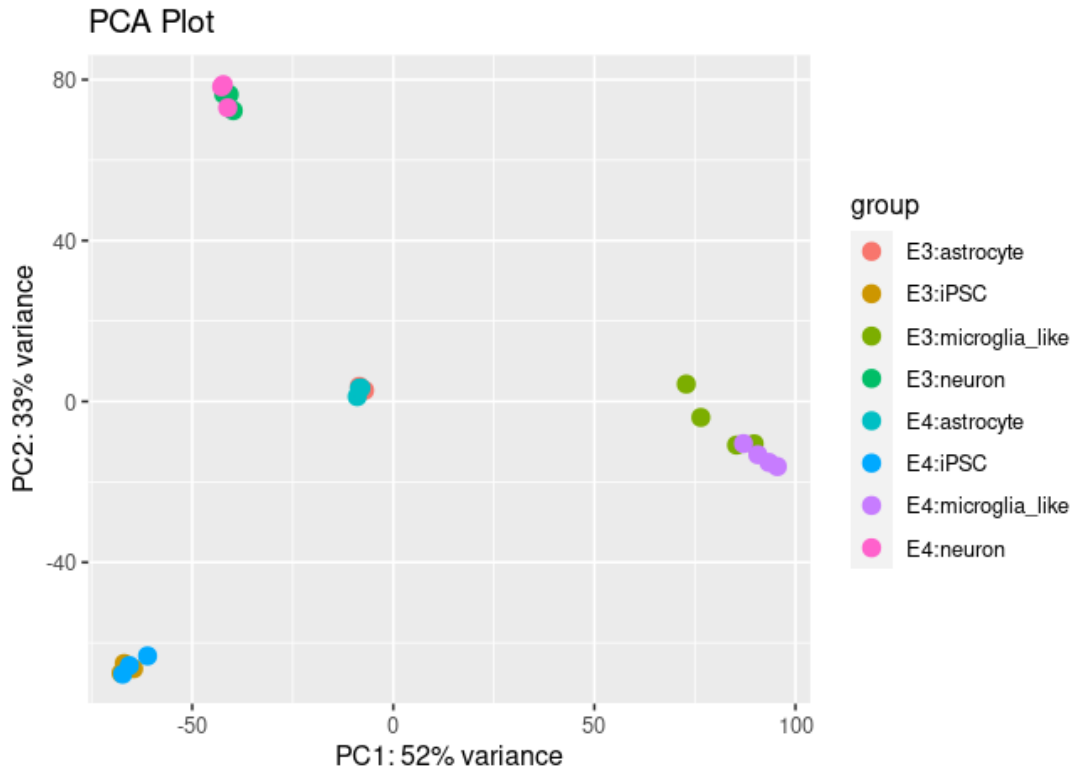

Figure 57 - PCA plot of PRJNA399530. Samples are clustered by cell types and are not separated by APOE variant.

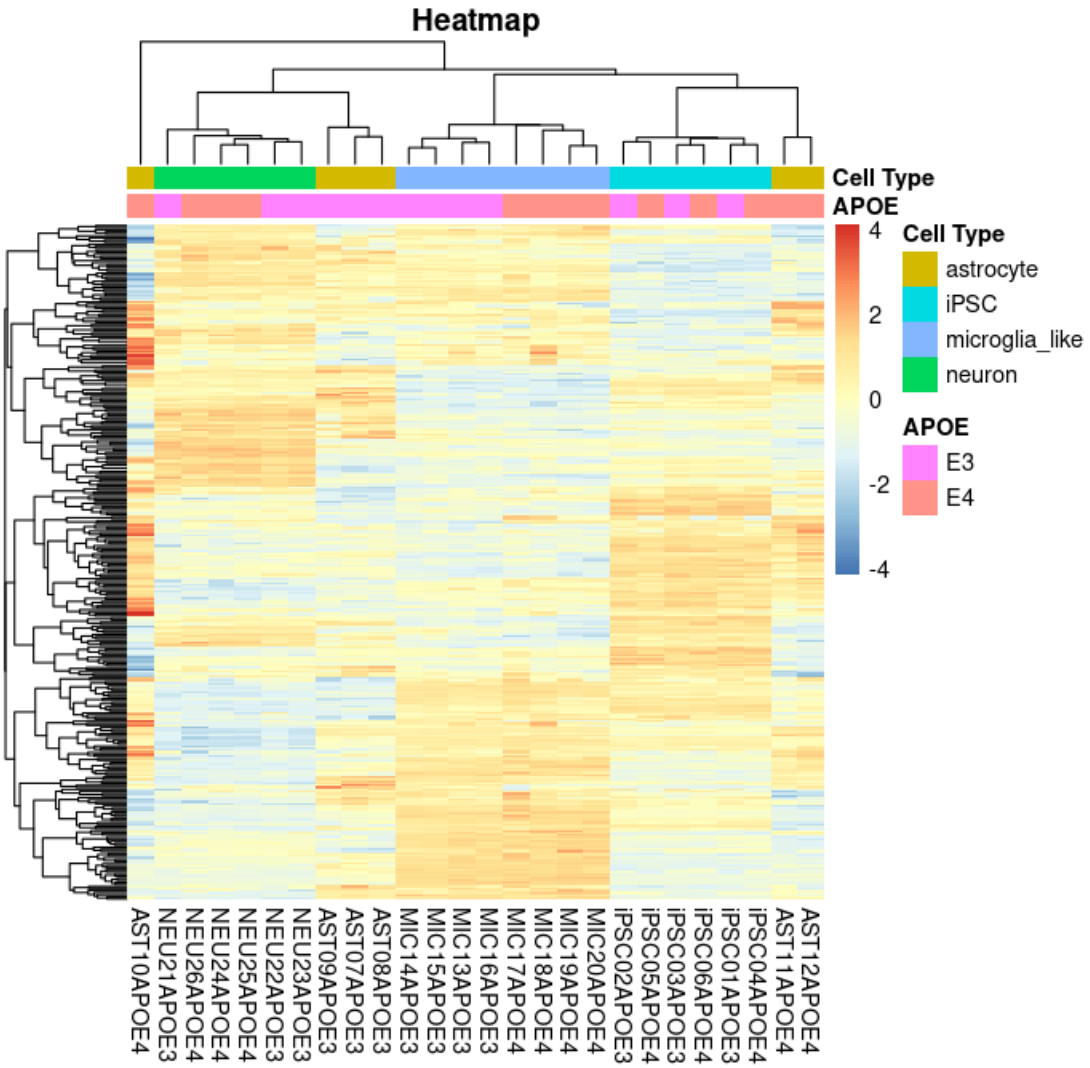

Figure 58 - Heatmap of PRJNA399530. Samples show a visible distinction in terms of significant gene expressions. Separation for APOE variant is not clear.

8.4. PRJNA451437

This dataset compares healthy and AD postmortem olfactory bulb samples.

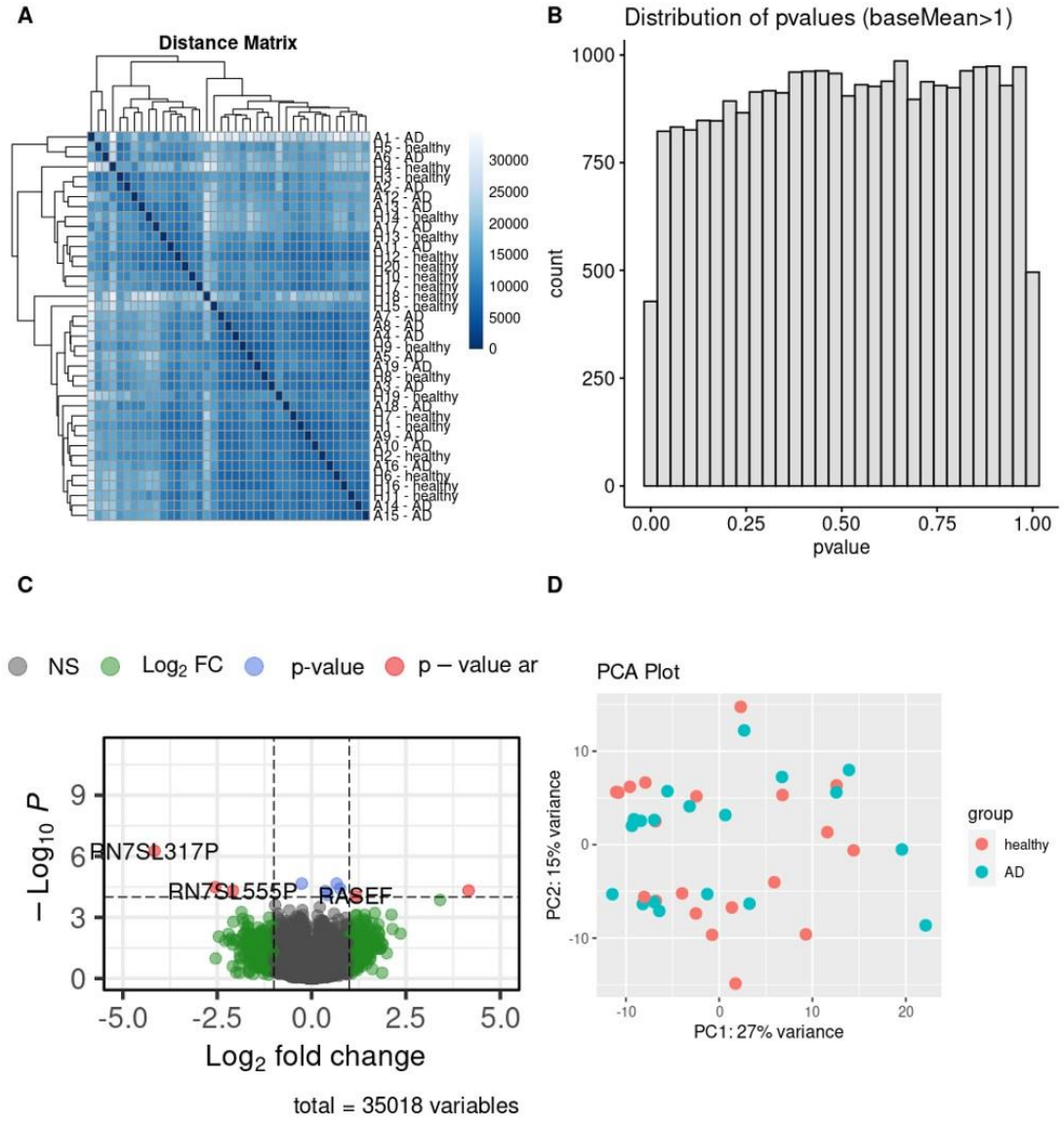

Figure 59 – Summary graphs of PRJNA451437. The genes have low significance scores and samples are not separated well. A heatmap cannot be drawn. A: Distance matrix. B: Distribution of p-values C: Volcano plot D: PCA Plot

8.5. PRJNA482601

This dataset compares healthy (APOE3) and AD-like (APOE4) iPSC-derived cerebral organoids.

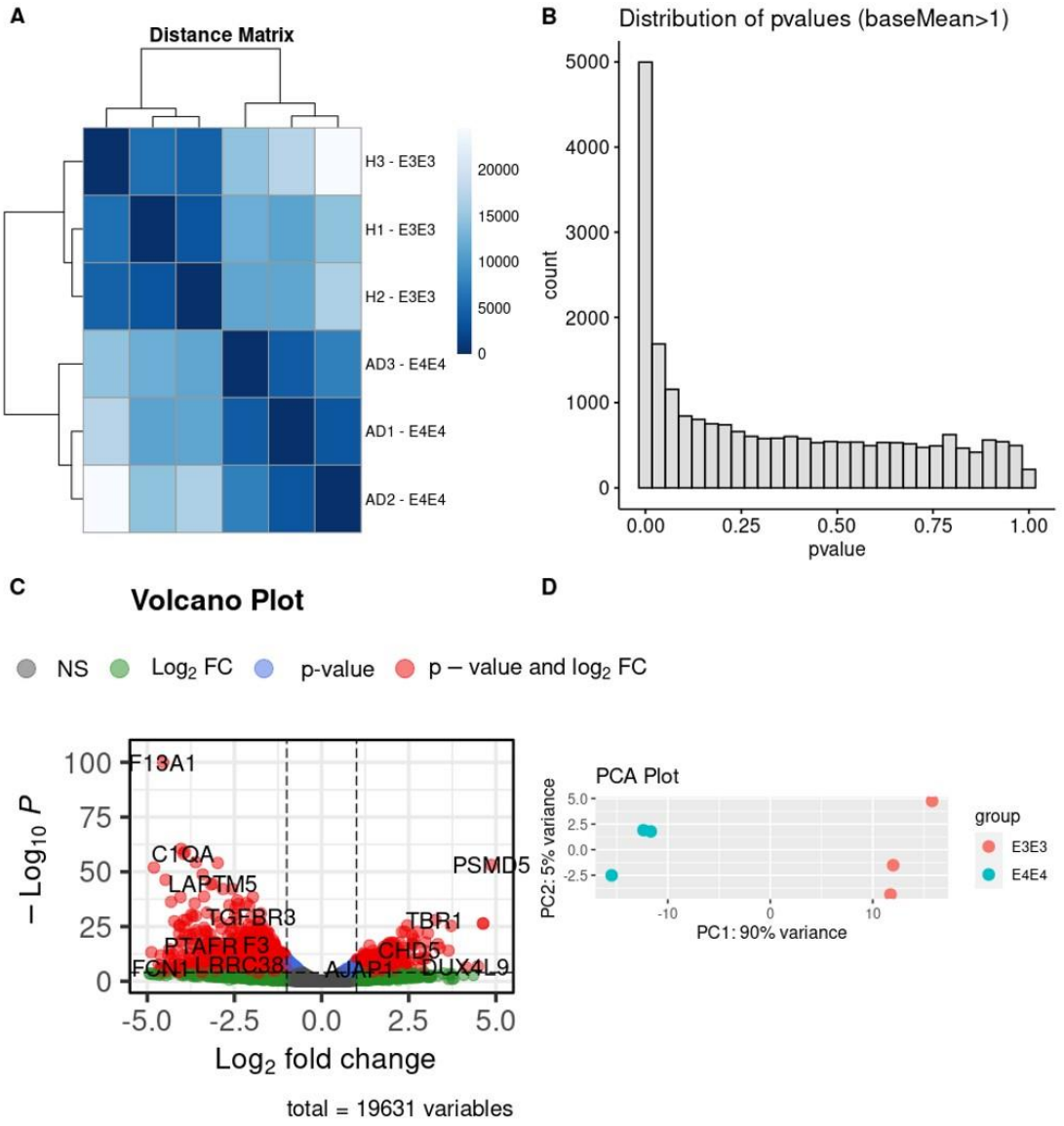

Figure 60 – Summary graphs of PRJNA482601. The samples have high amount of significant genes and they are separated well. On both the distance matrix and the PCA plot, the samples are clearly separated by their APOE type. A: Distance matrix. B: Distribution of p-values. C: Volcano plot D: PCA Plot

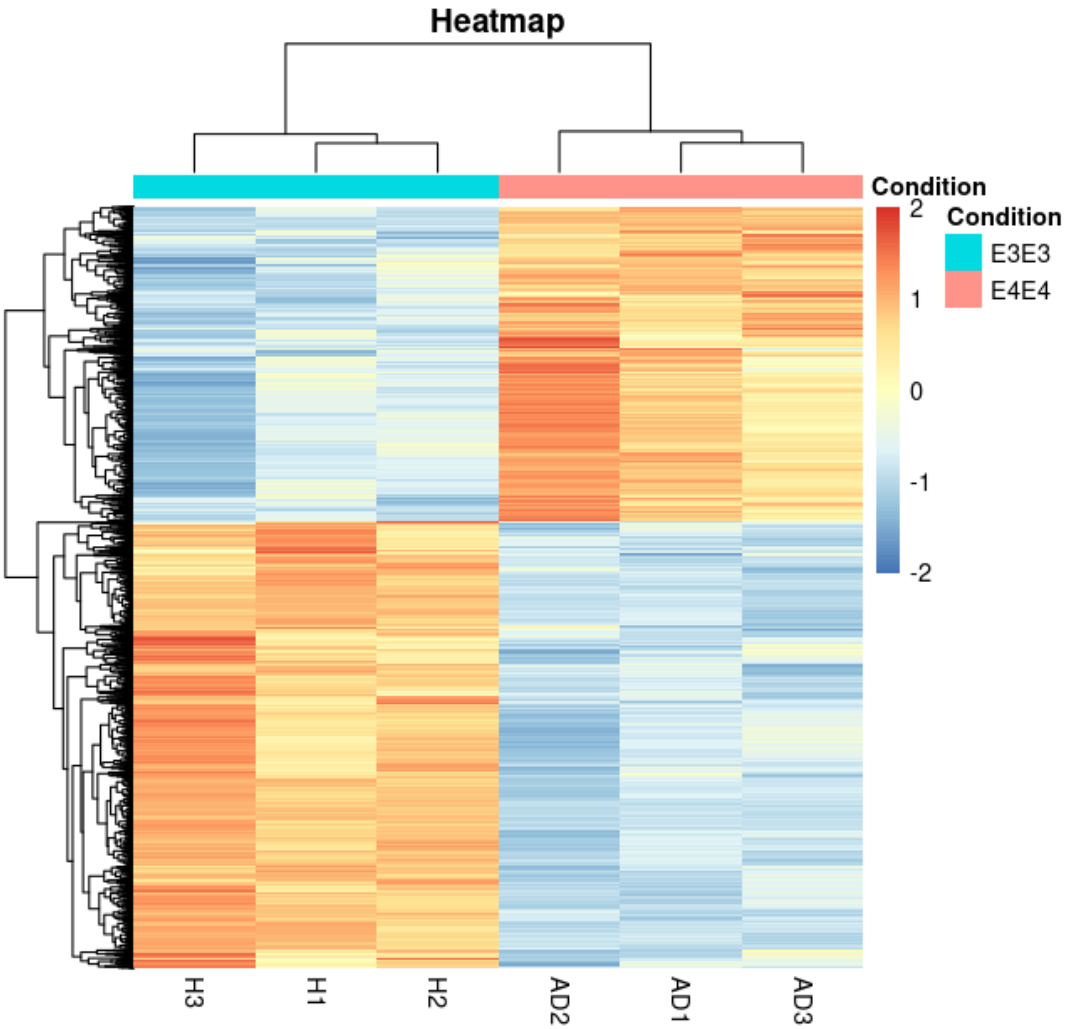

Figure 61 - Heatmap of PRJNA482601. There is a distinct expression profile between AD-like and healthy samples.

8.6. PRJNA527202

This dataset compares healthy (WT) and AD-like (APP, PSEN1 and APP+PSEN1 mutant) iPSC-derived neurons.

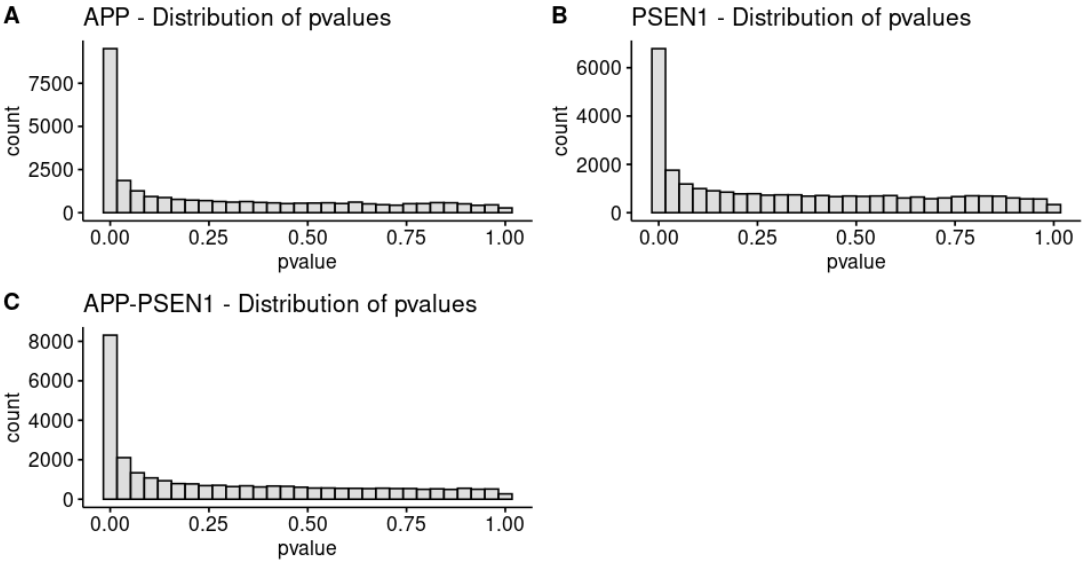

Figure 62 – P-value distributions of PRJNA527202. Each sample carry high amount of significant genes.

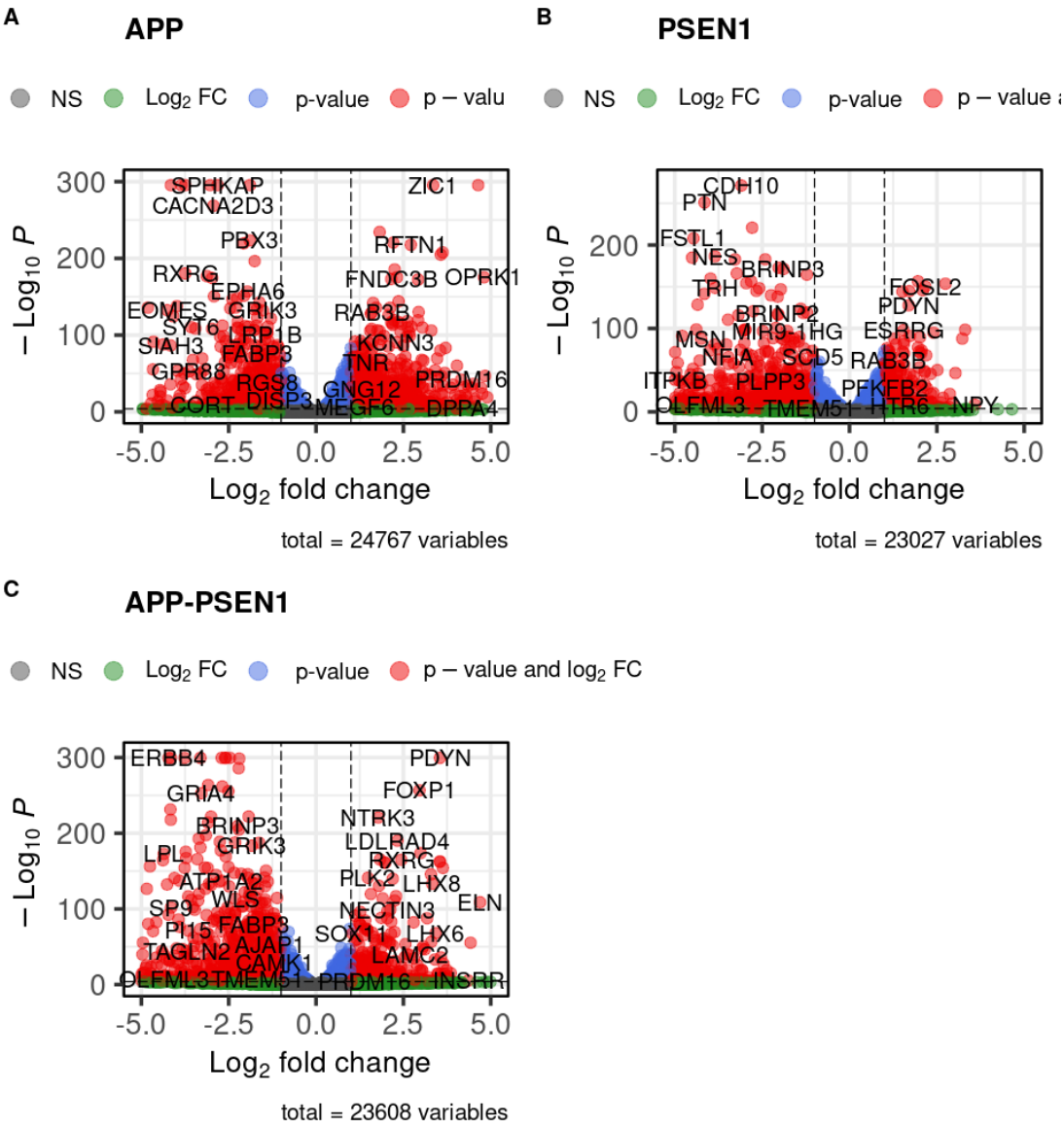

Figure 63 - Volcano plots of PRJNA527202. As seen with the red dots, there are high amount of DEGs.

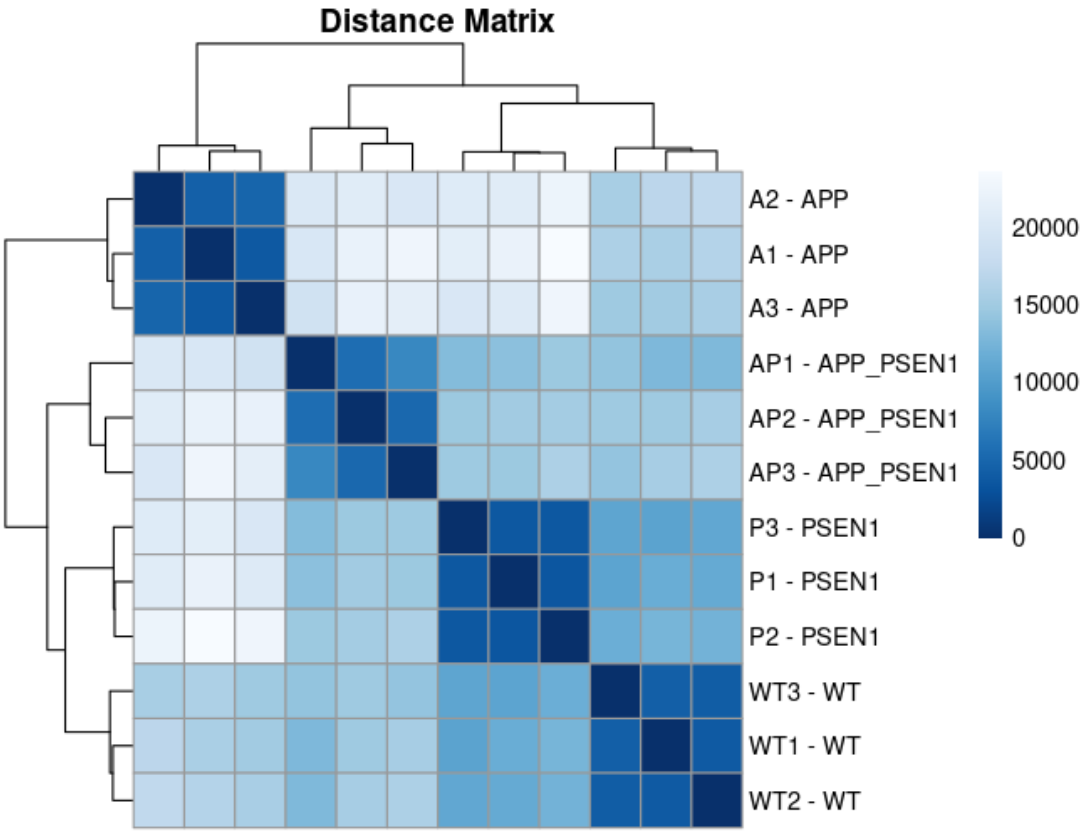

Figure 64 - Distance matrix of PRJNA527202. Samples show a clear separation by their mutation.

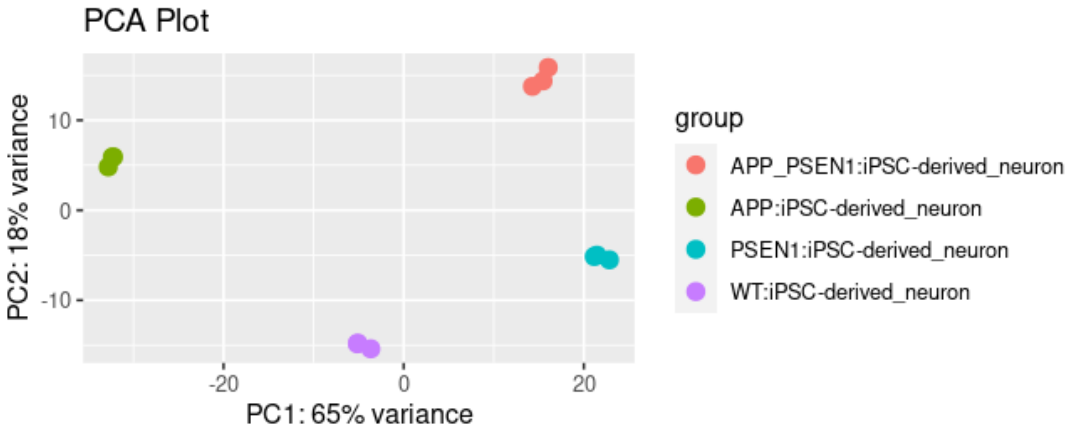

Figure 65 - PCA plot of PRJNA527202. Samples are separated by their mutation type with very distinct clusters.

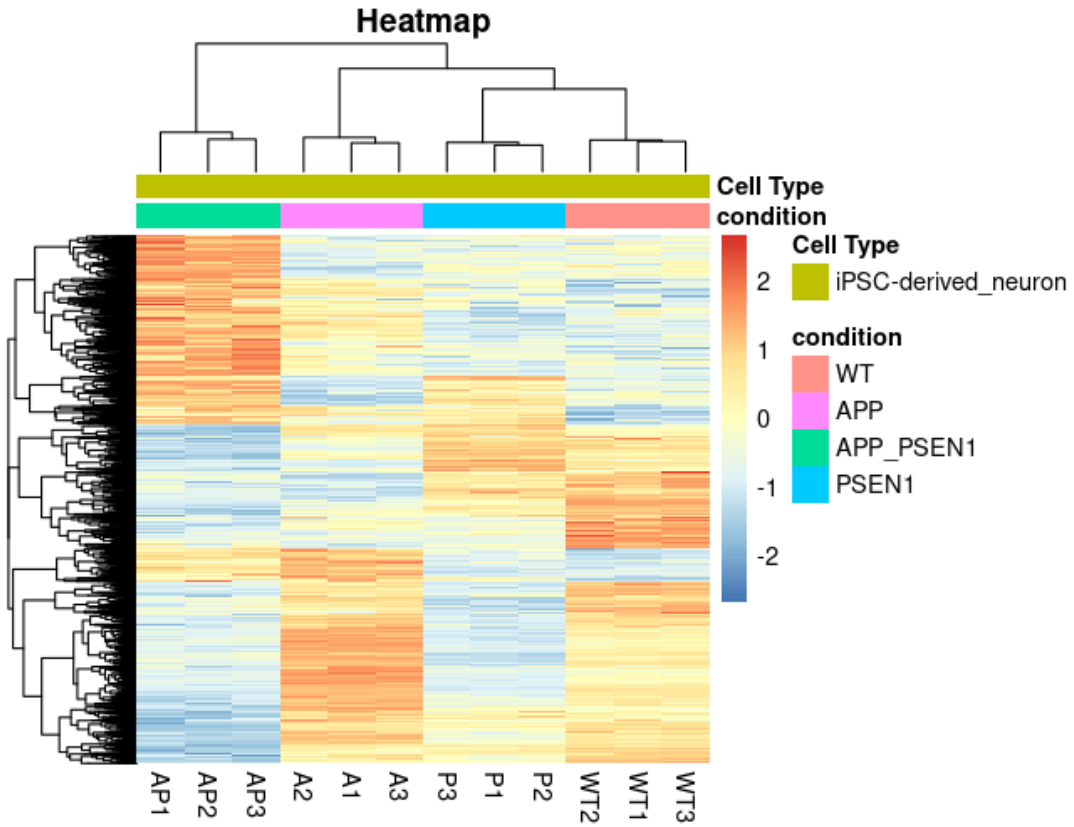

Figure 66 - Heatmap of PRJNA527202. Samples show a very distinct expression profile by their mutation type.

8.7. PRJNA559812

This dataset compares healthy (PSEN1-deltaE9 corrected) and AD-like (PSEN1-deltaE9) iPSC-derived microglia samples.

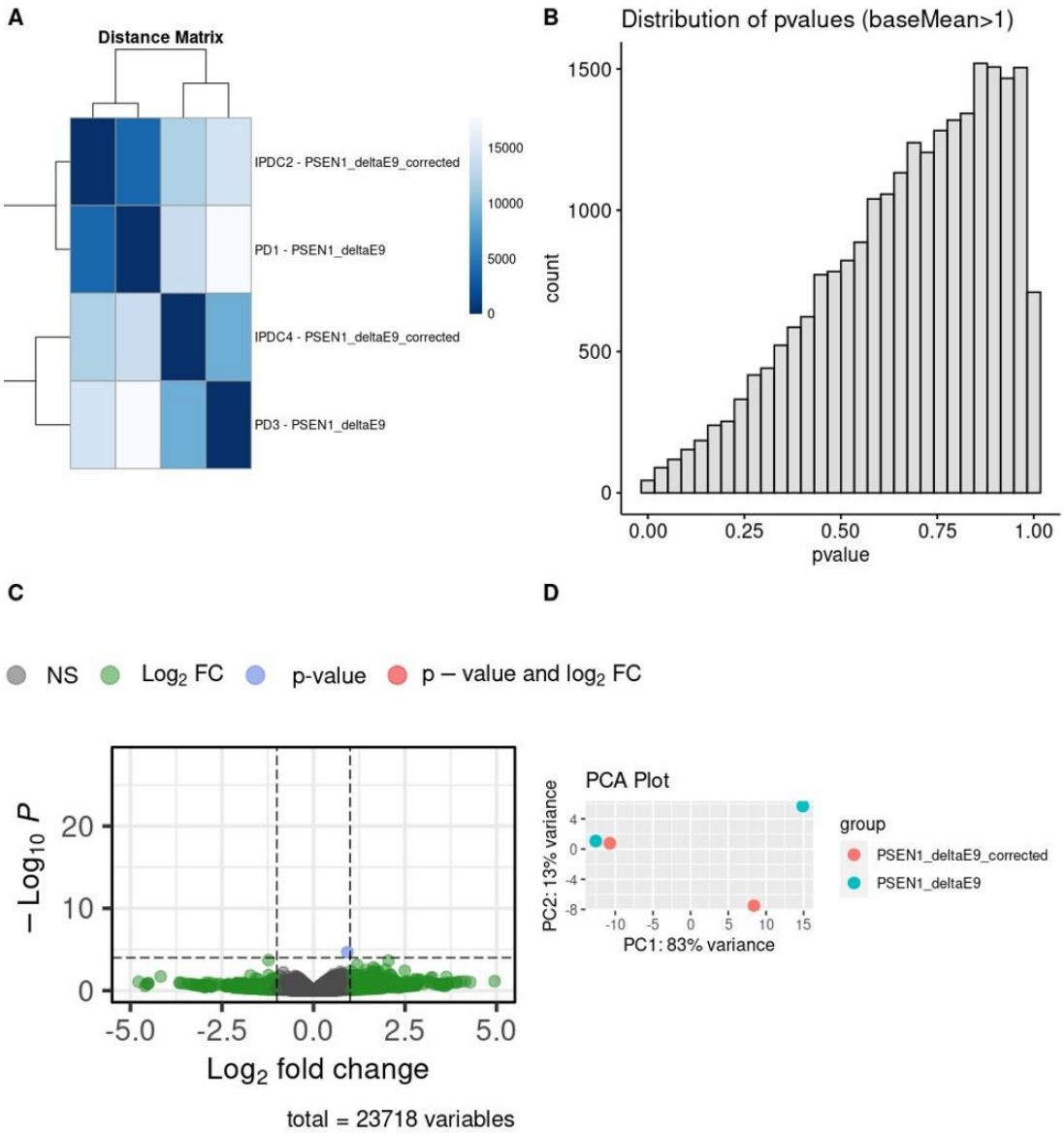

Figure 67 - Summary graphs of PRJNA559812. The samples are not separated by their PSEN1 type and there are no significant genes.

8.8. PRJNA576835

This dataset compares healthy (PSEN1-deltaE9 corrected) and AD-like (PSEN1-deltaE9) iPSC-derived astrocyte samples.

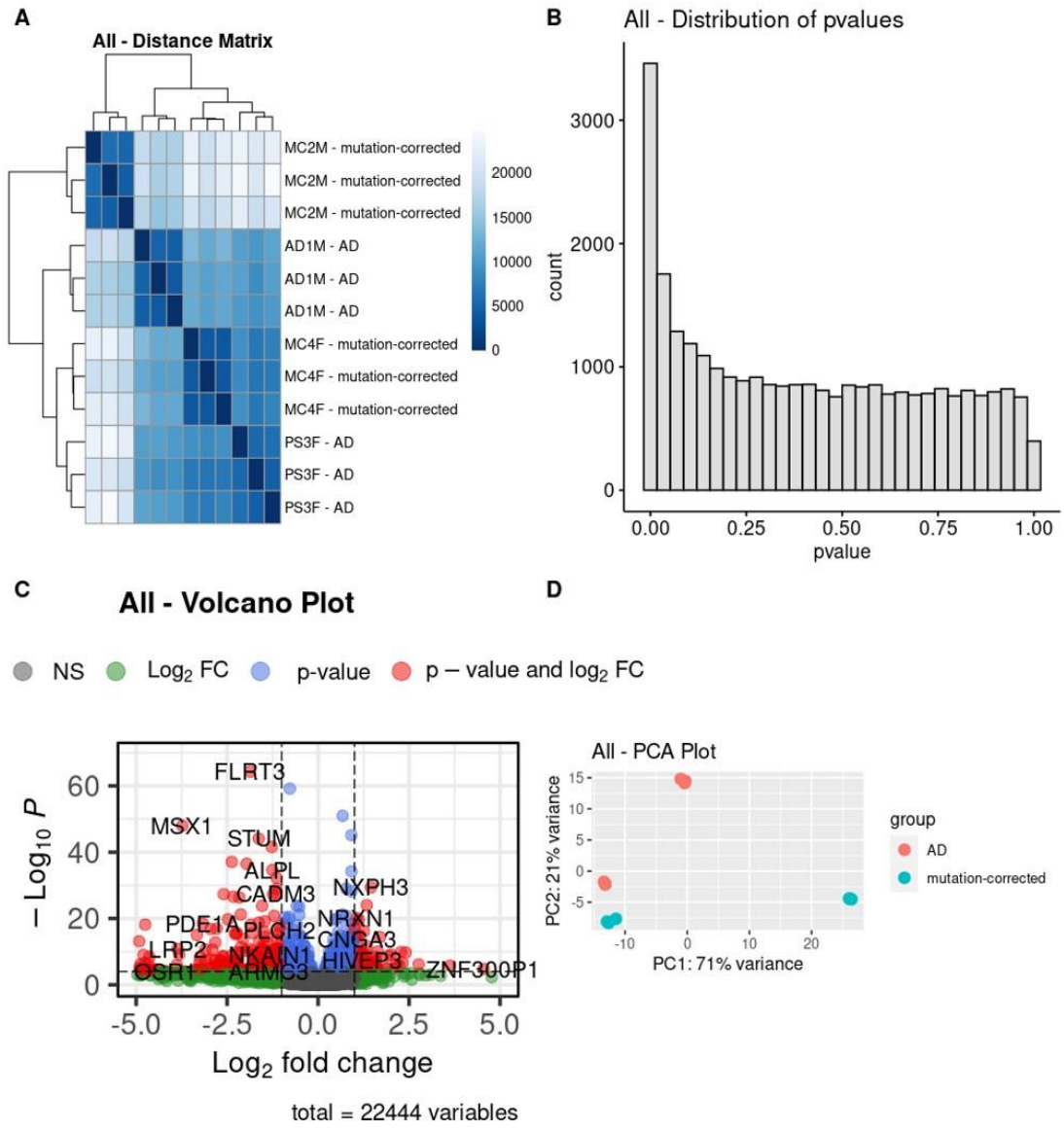

Figure 68 - Summary graphs of PRJNA576835. There are goo amount of significant genes and the samples are separated by their PSEN1 variant to an extent. DEGs are mostly underexpressed. Sex is not considered in these comparisons.

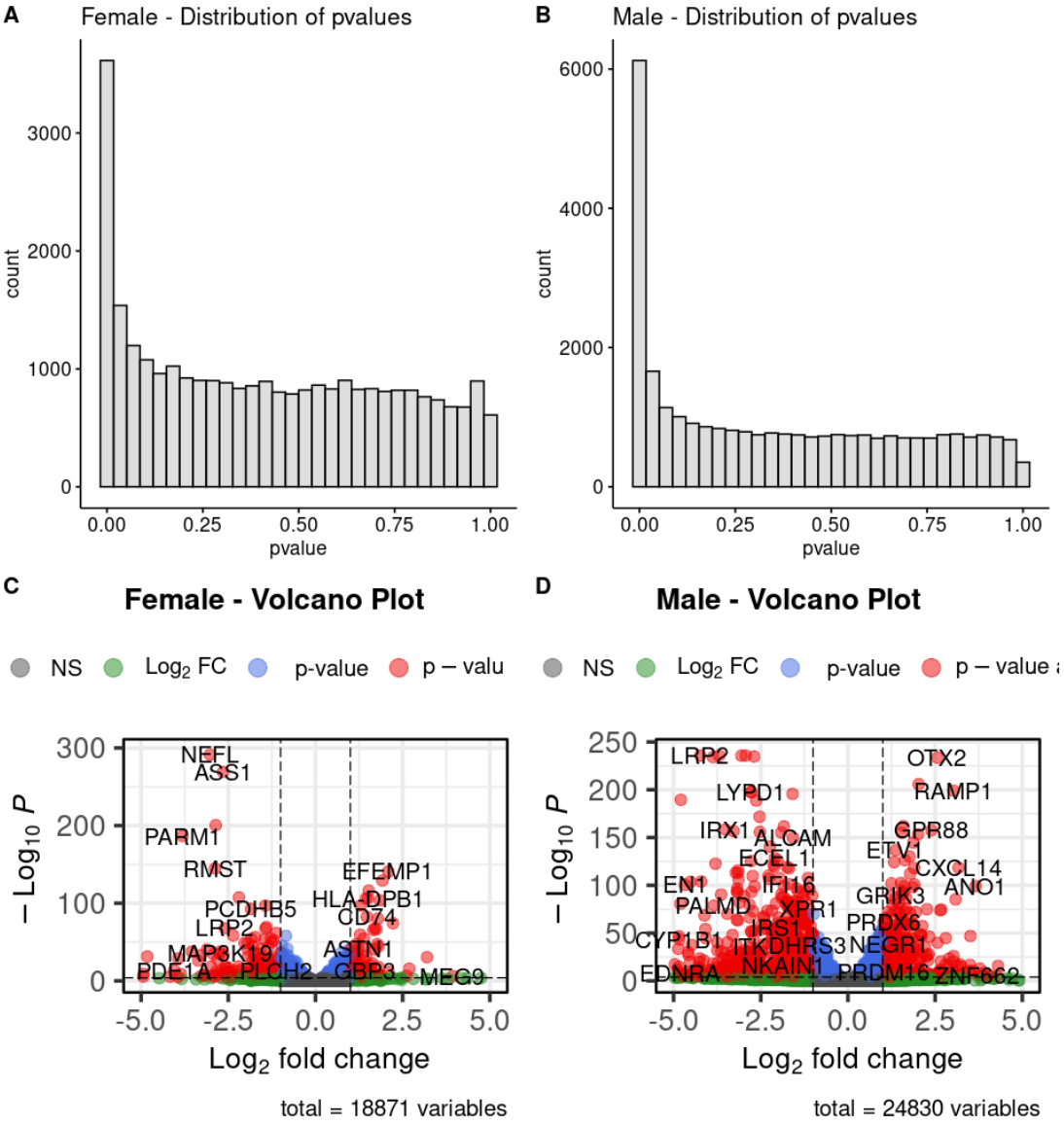

Figure 69 - P-value distributions and volcano plots by sex. Male samples have higher amount of significant genes.

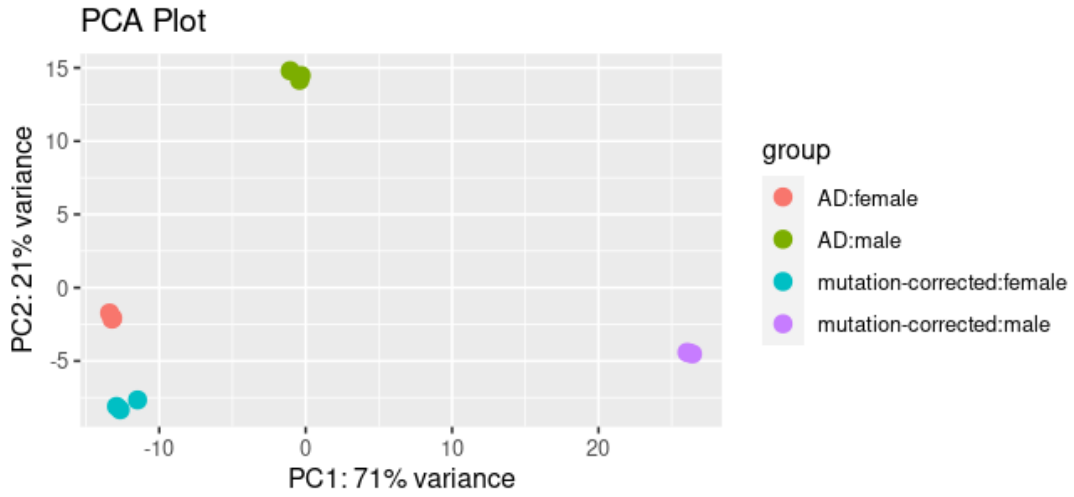

Figure 70 - PCA plot of PRJNA576835. Samples are distinctly separated by sex and condition. Male samples are more distant than female samples, as they have more significant genes.

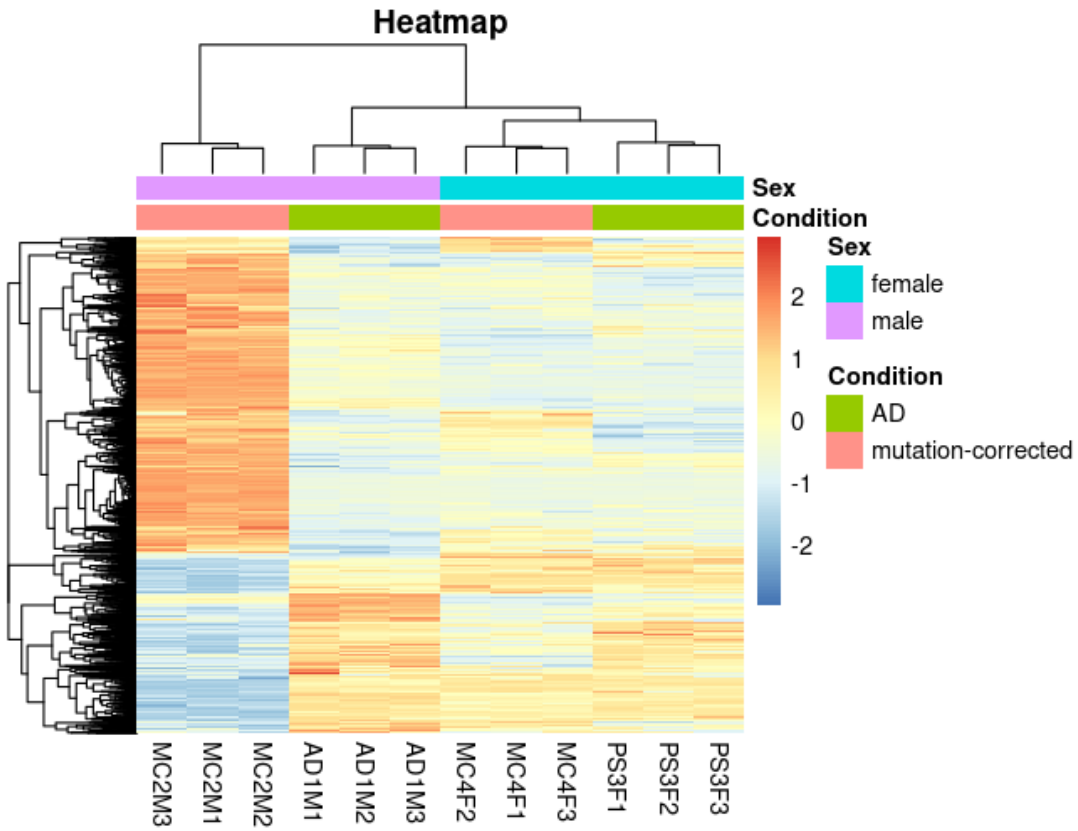

Figure 71 - Heatmap of PRJNA576835. Expression profile of male controls are more distinct than others.

8.9. PRJNA603192

This dataset compares female and male samples by their APOE type (APOE2, APOE3, APOE4) from inferior parietal lobule postmortem tissues.

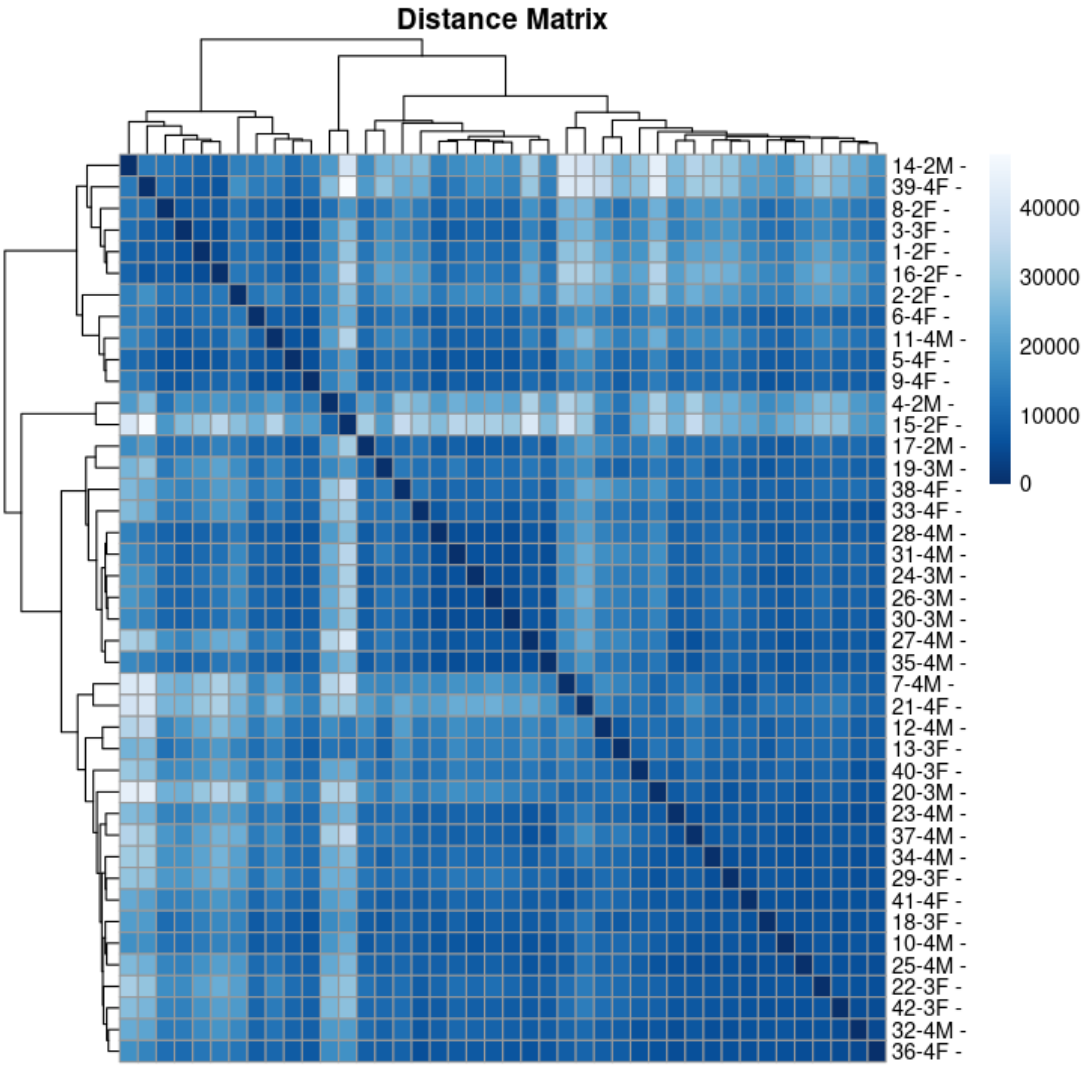

Figure 72 - Distance matrix of PRJNA603192. APOE2 samples are clustered together to an extent while APOE3 and APOE4 samples are mixed.

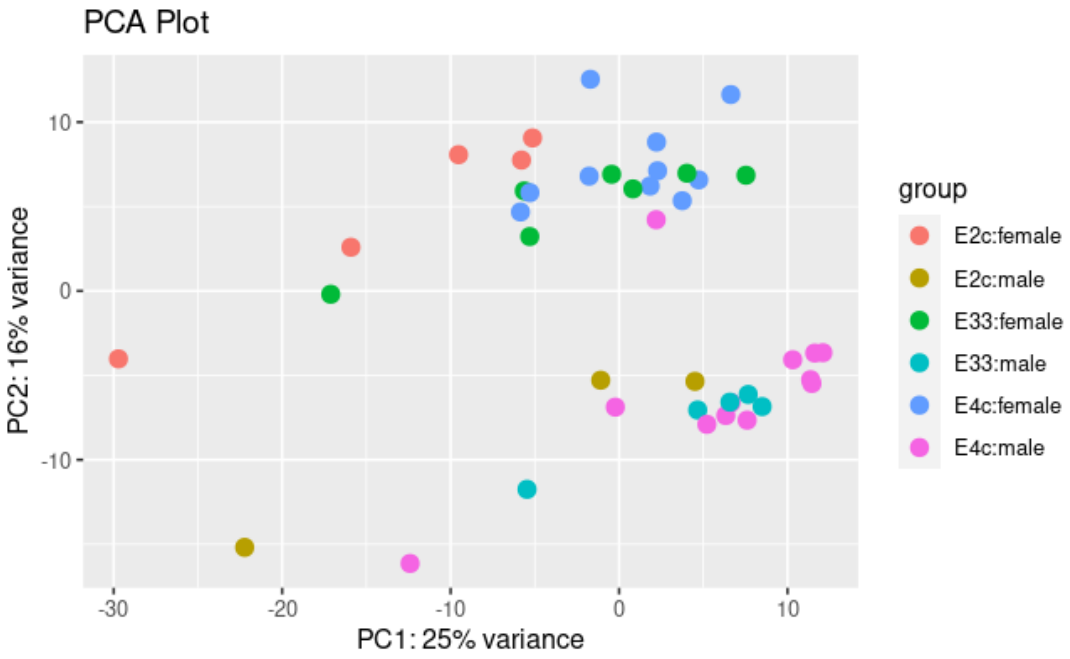

Figure 73 - PCA plot of PRJNA603192. PC2 separates sexes clearly. Female clusters are observed on the top while male clusters are on the bottom side. PC1 separates the samples by their condition. E2 samples lean toward left, while E3 and E4 samples are mixed and lean towards right.

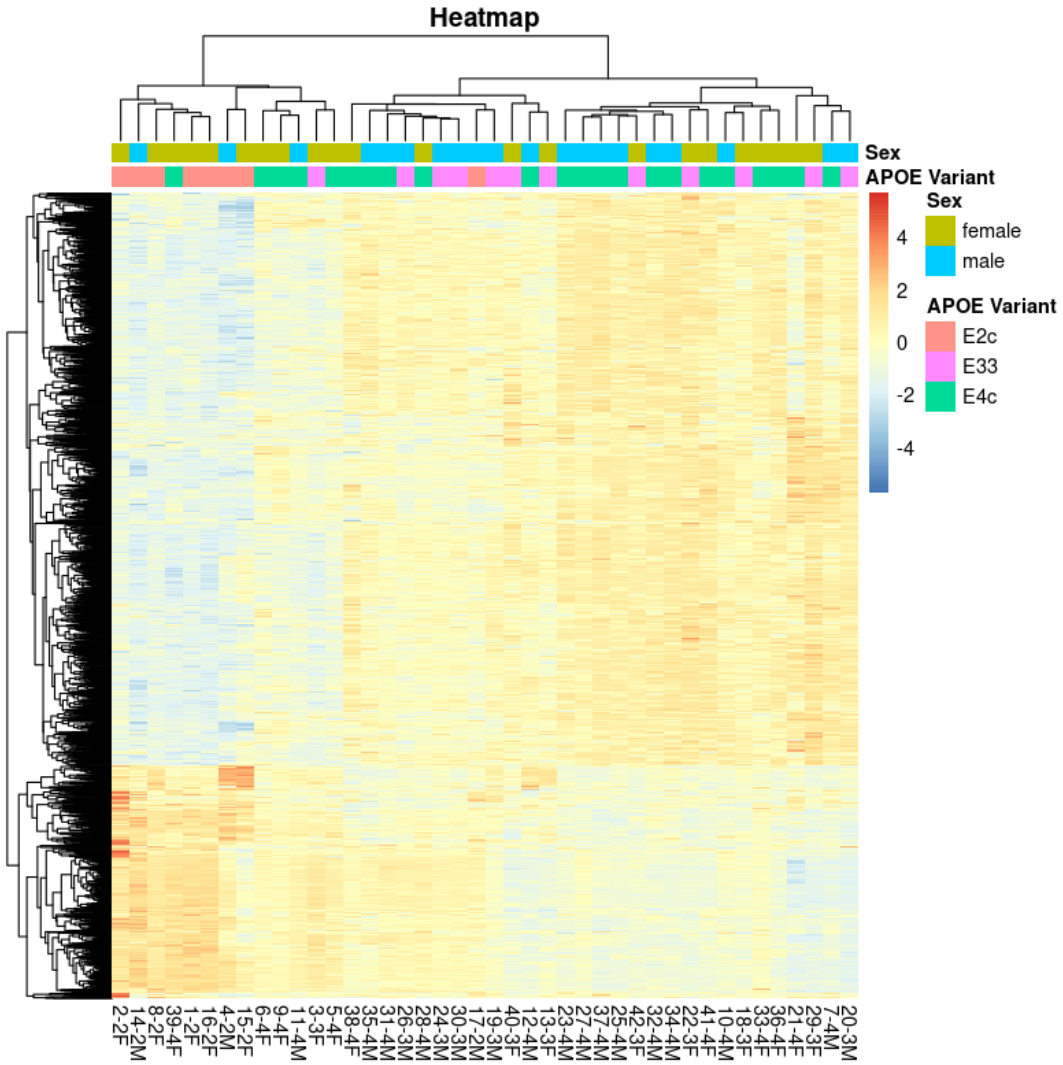

Figure 74 - Heatmap of PRJAN603192. Female APOE2 carriers are positioned on the left. The rightmost side contains mostly female APOE4 carriers. Leftmost and rightmost sides show the most distinct expression profile.

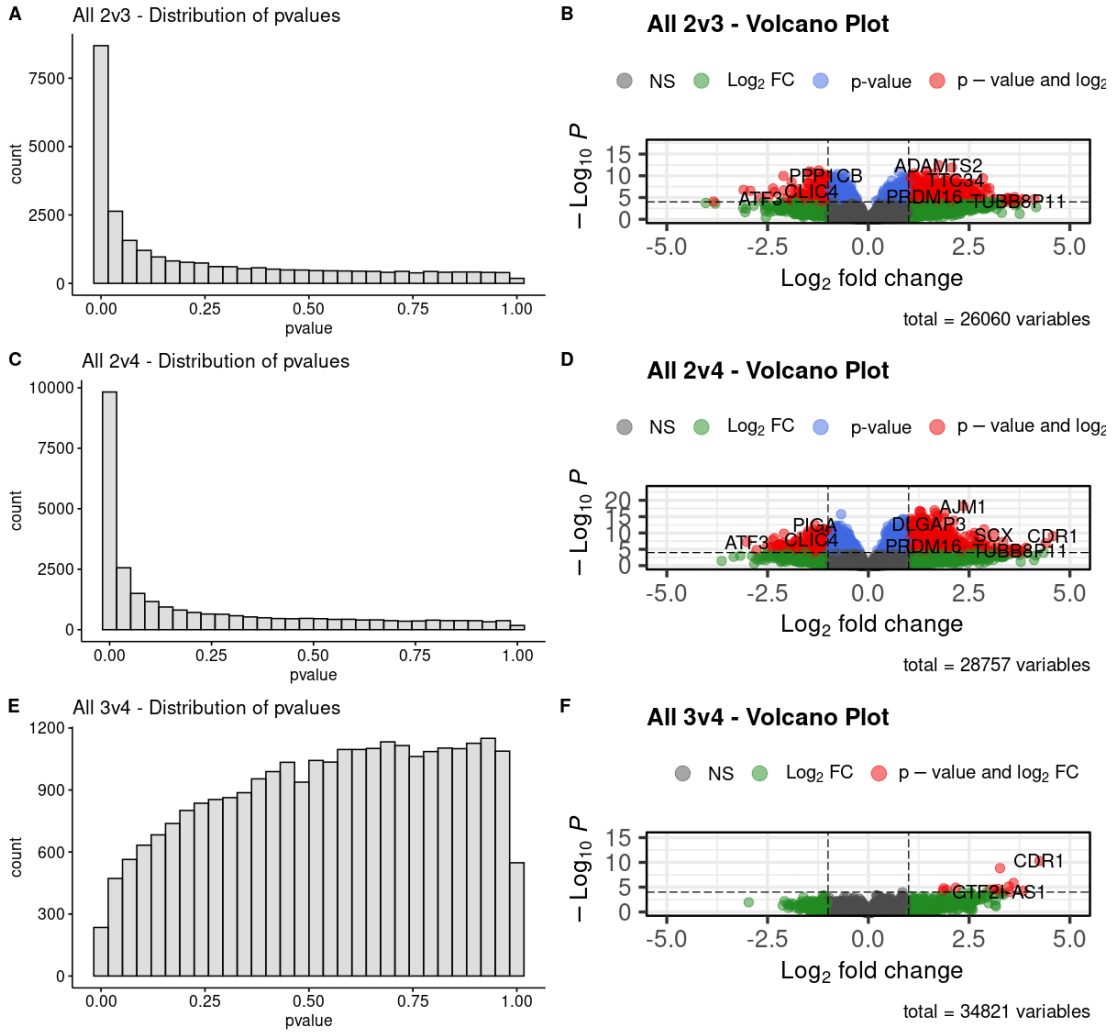

Figure 75 - P-value distributions and volcano plots of samples without considering sex variable. APOE3-APOE4 comparison doesn't provide DEGs, in contrast to other comparisons.

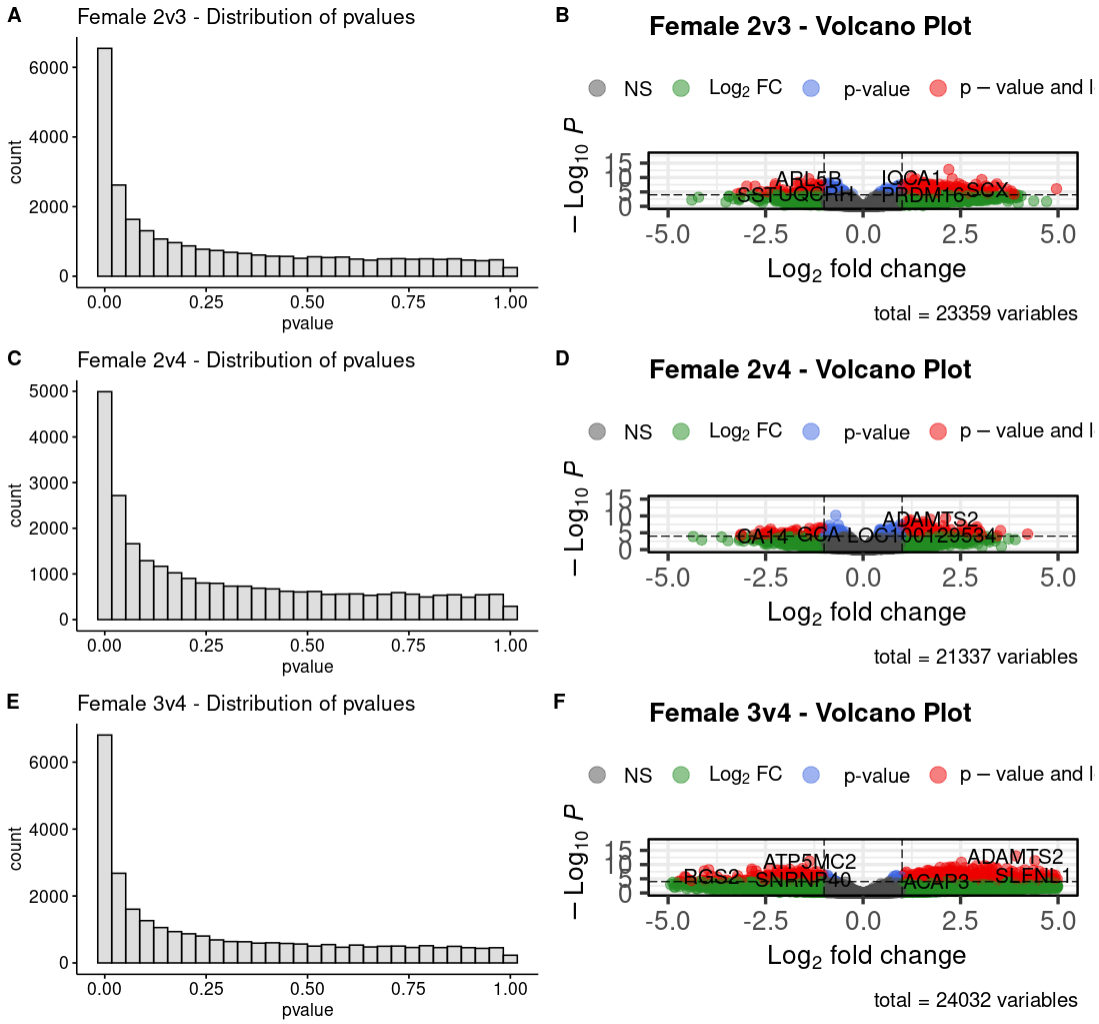

Figure 76 - P-value distributions and volcano plots of female samples. All comparisons show high amount of DEGs.

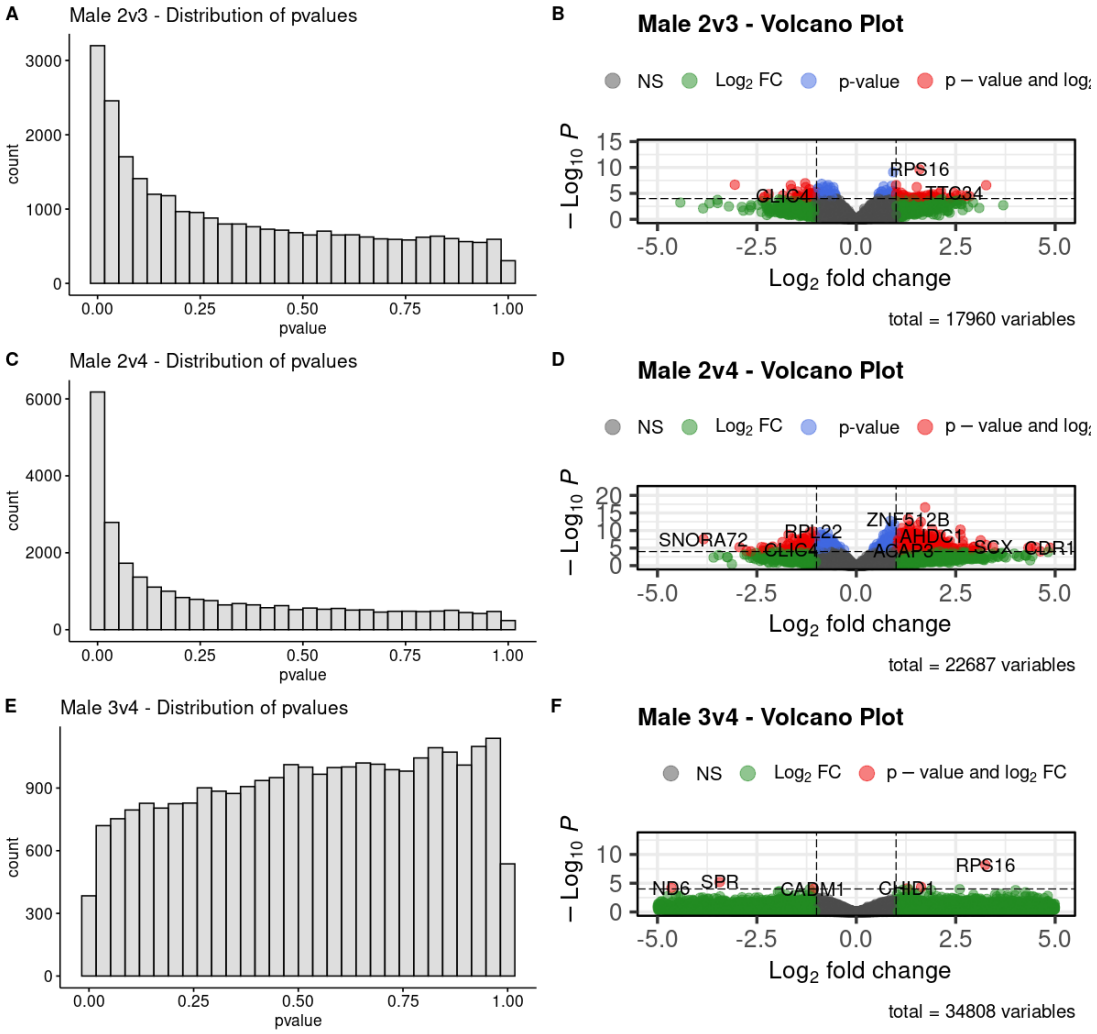

Figure 77 - P-value distributions and volcano plots of male samples. APOE3-APOE4 comparison doesn't show any significant genes in contrast to other comparisons. Still, the number of significant genes are low relative to the female samples.

8.10. PRJNA643561

This dataset compares healthy (WT) and AD-like (SORL1-KO, SORL1-variant, TREM2-KO, TREM2-variant) iPSC-derived microglia samples.

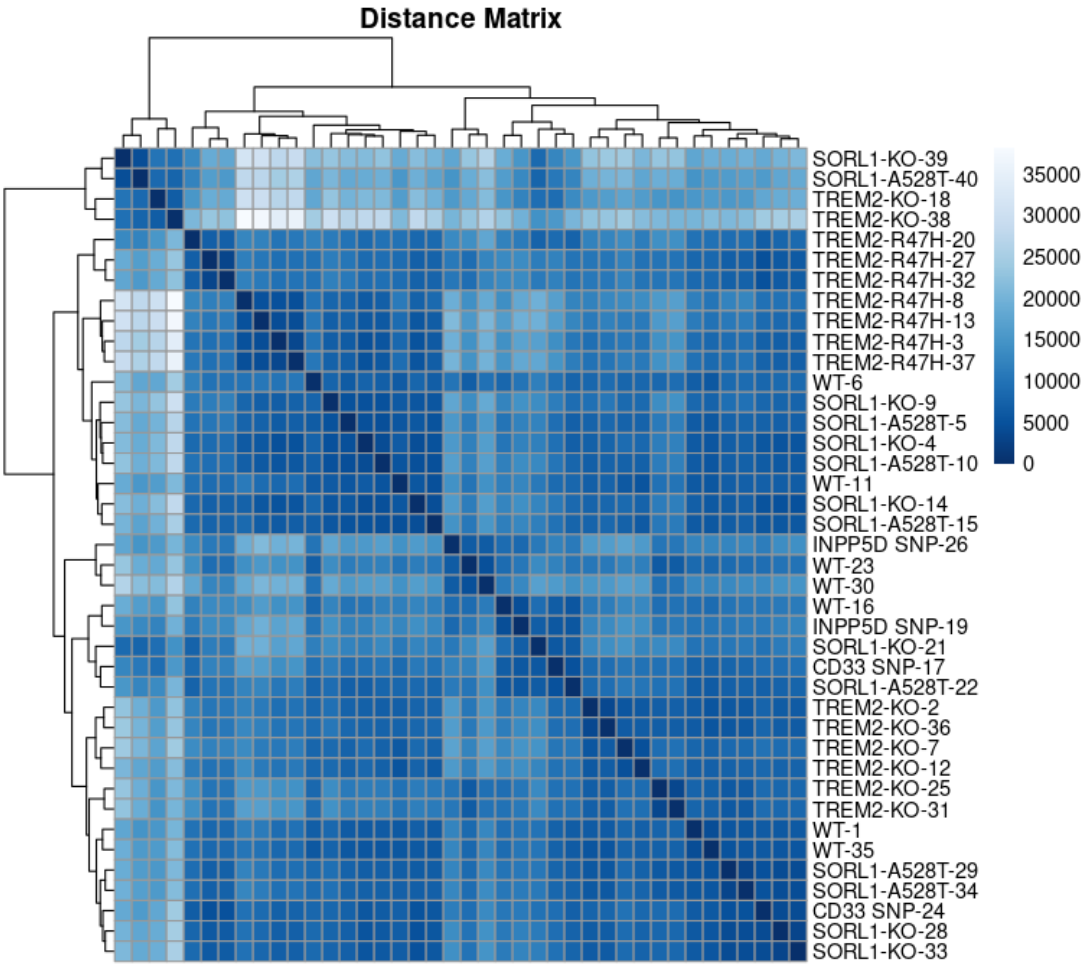

Figure 78 - Distance matrix of PRJNA643561. TREM2 mutant are clustered together by being a variant or KO while SORL1 clusters are mixed together.

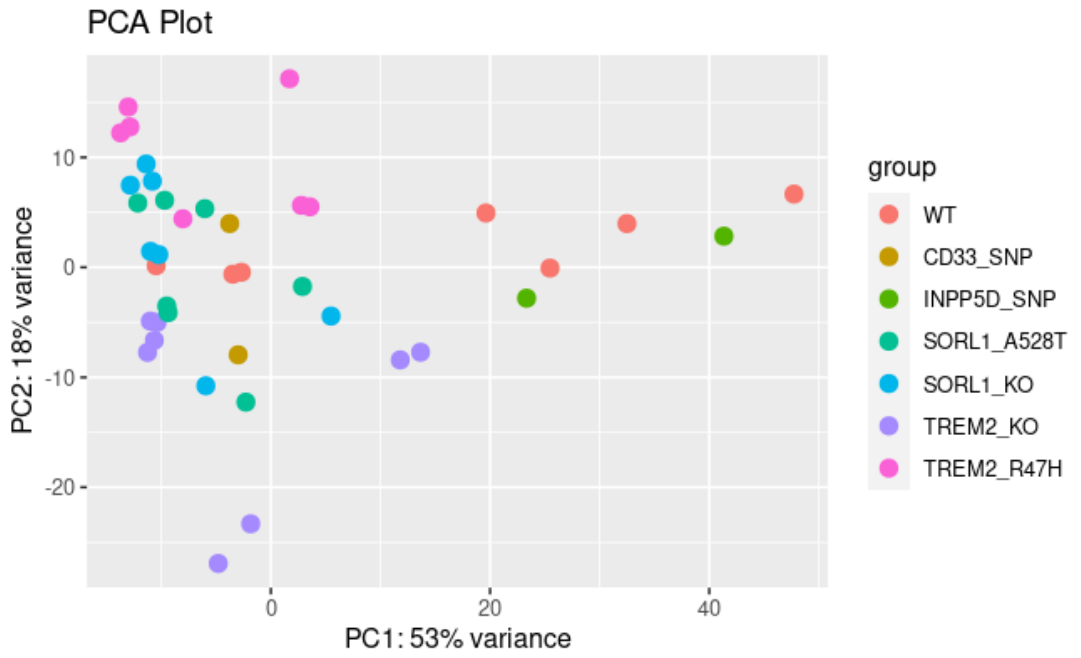

Figure 79 – PCA plot of PRJNA643561. PC1 separates the samples by their conditon: WT samples are leaning towards right while all other samples are towards left. PC2 separates TREM2-KO samples from other, mostly.

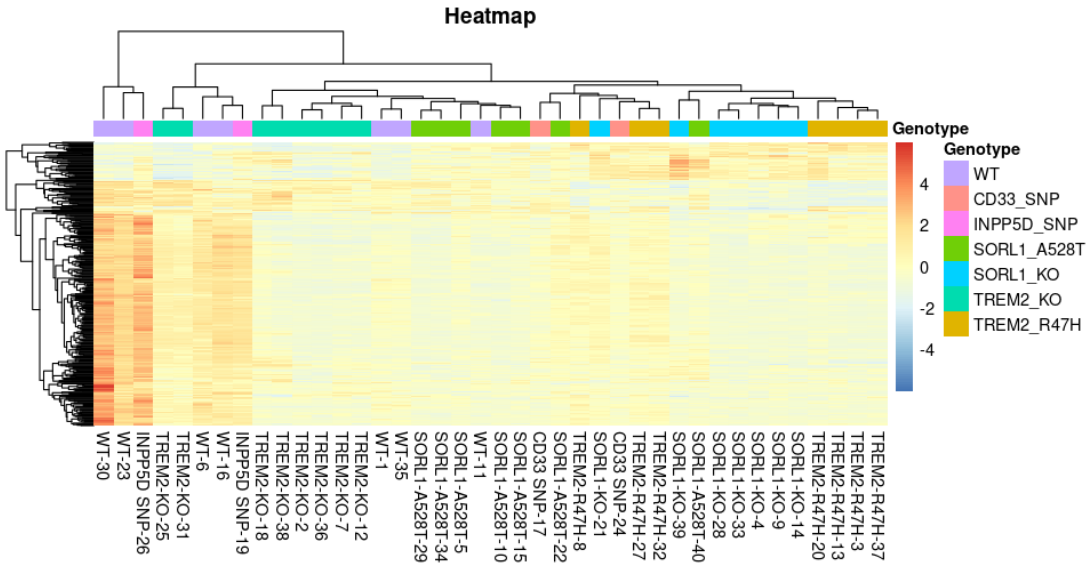

Figure 80 - Heatmap of PRJNA643561. WT and TREM2-KO show distinct profiles. The most distinct samples are WT samples.

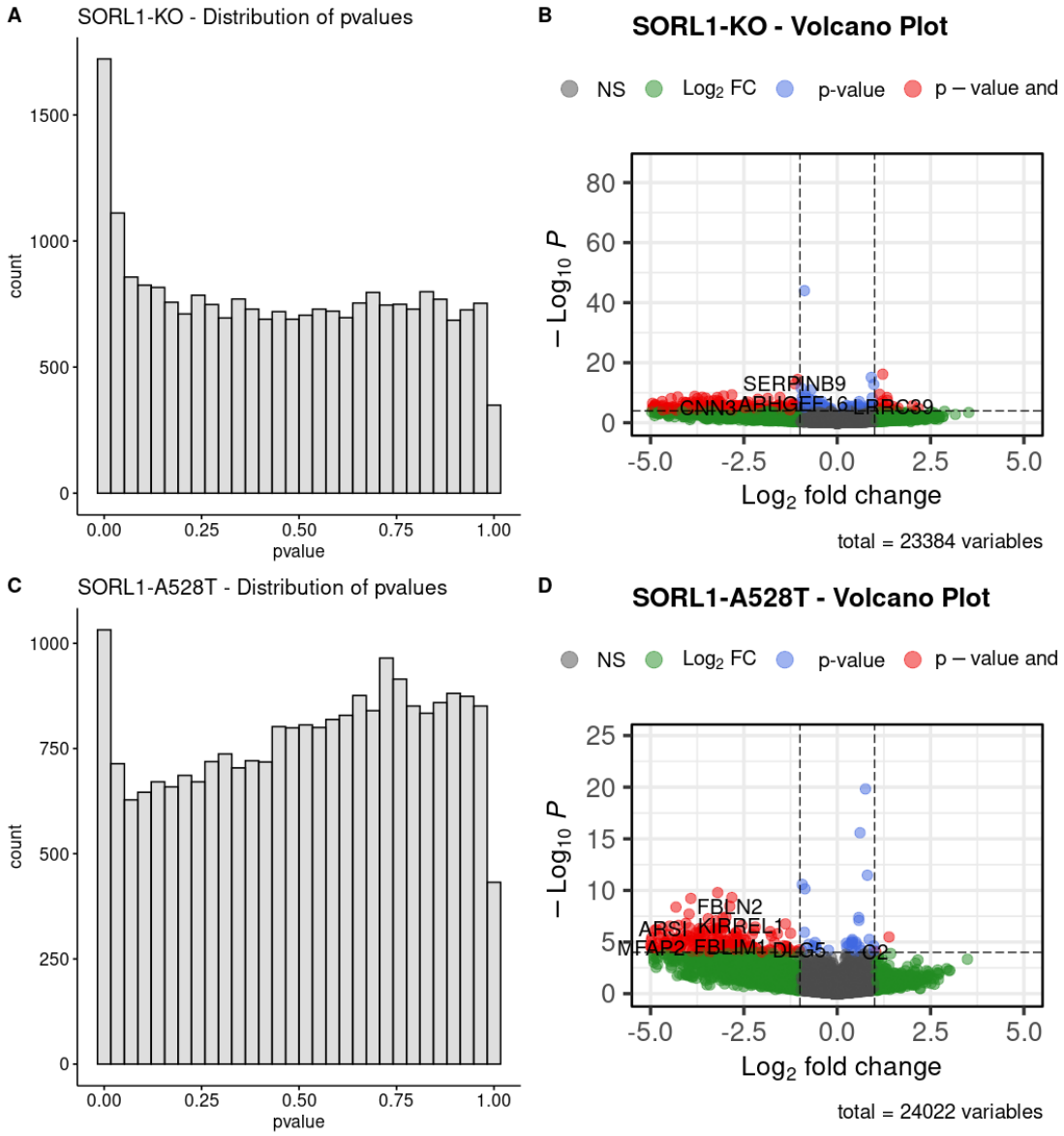

Figure 81 - P-value distribution and volcano plots of SORL1 mutants. KO mutant are showing more significant genes. Both of the samples' significant genes are showing underexpression.

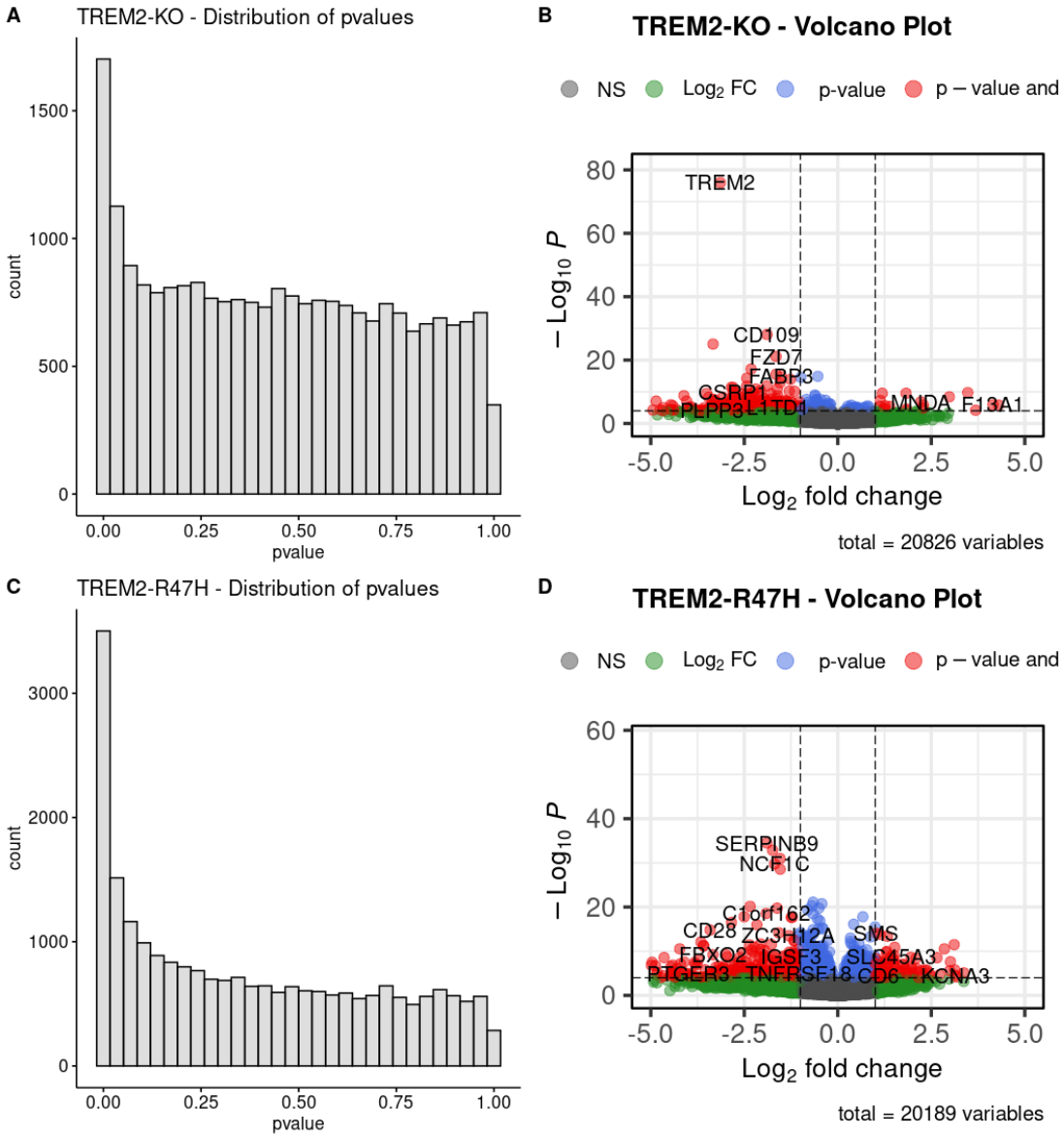

Figure 82 - P-value distributions and volcano plots of TREM2 mutants. TREM2-variant mutant is showing higher amount of DEGs. Most of the DEGs for both samples are showing underexpression.

8.11. PRJNA644383

This dataset compares healthy and AD postmortem lateral temporal lobe samples.

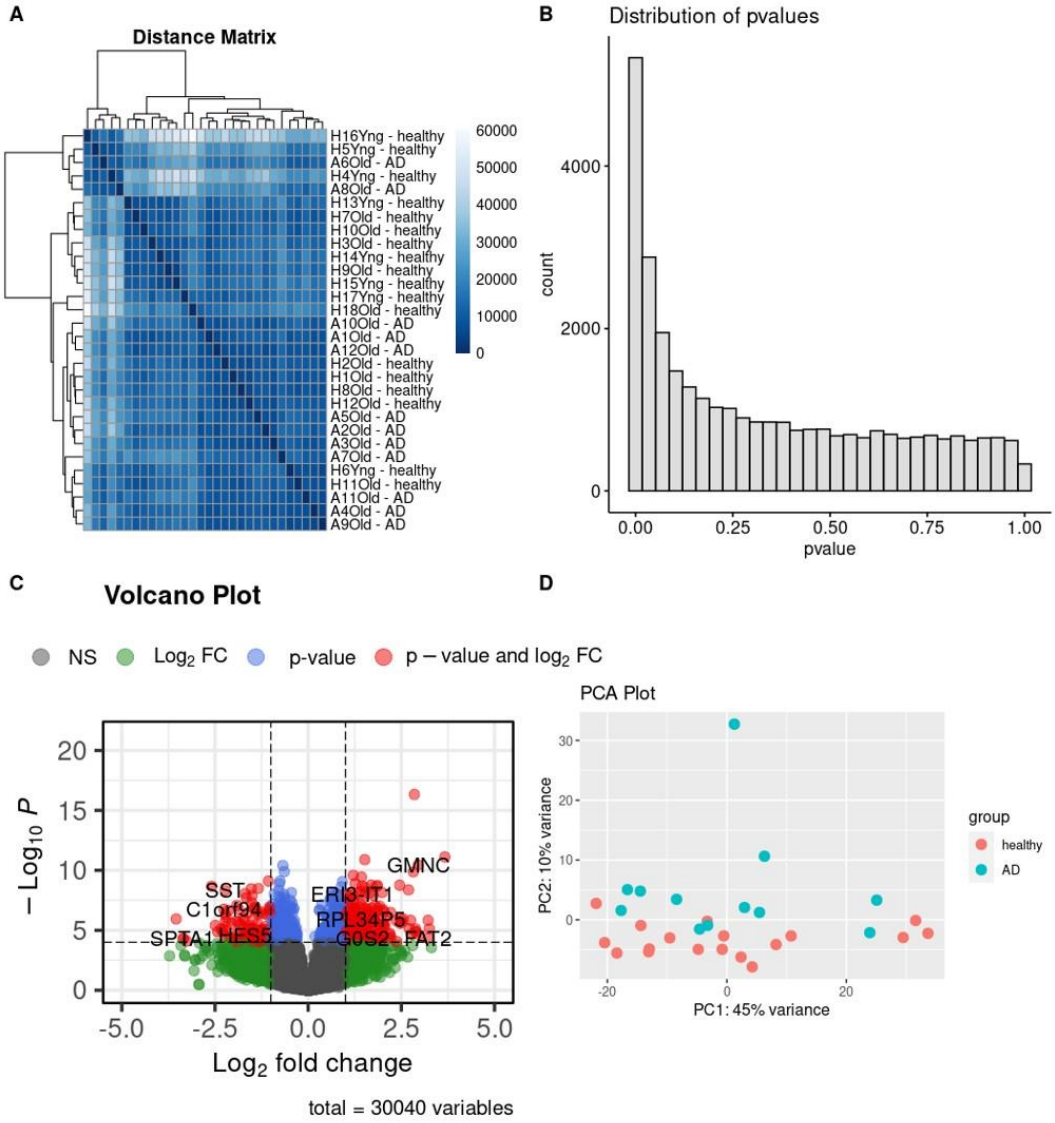

Figure 83 - Summary graphs of PRJNA644383. Distinction between healthy and AD samples is existant but in low amounts. There are high amount of significant genes, but their expression values are not in the extremes.

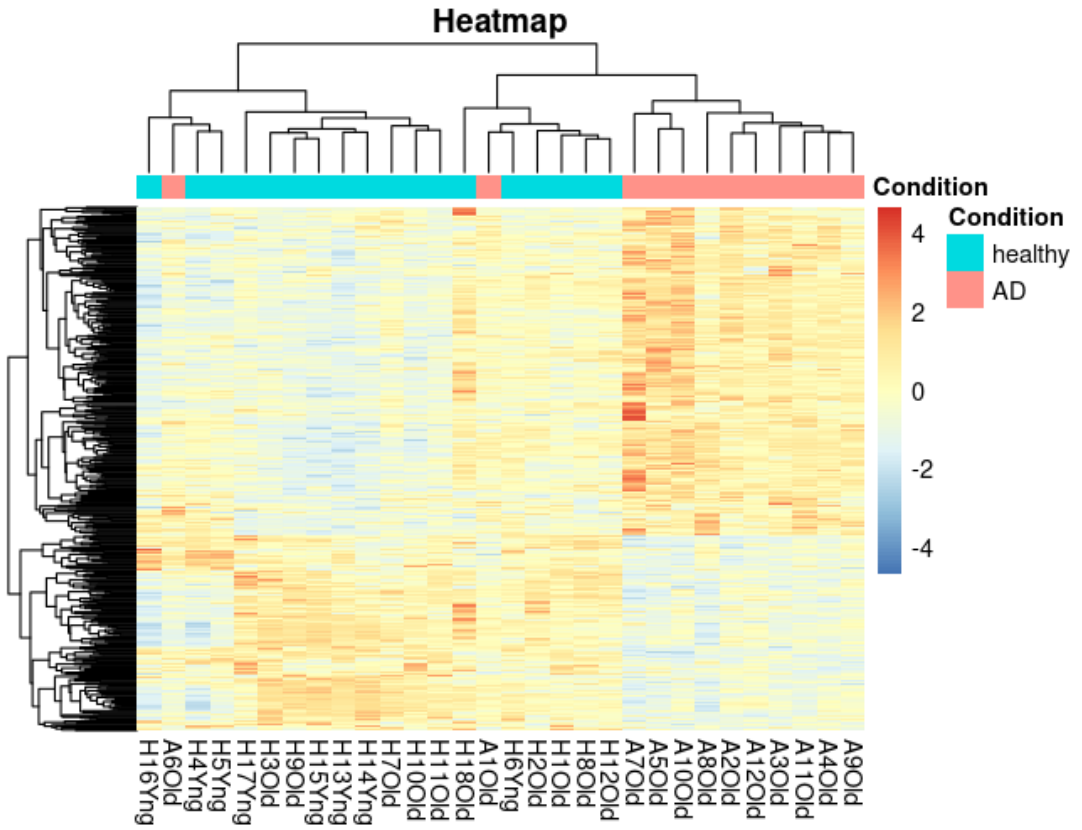

Figure 84 - Heatmap of PRJNA644383. AD and healthy samples are showing a very clear distinction in their expression profiles. There are only two AD samples showing similarities to healthy samples.

8.12. PRJNA662330

This dataset compares healthy (WT) and AD-like (TREM2-KO) iPSC-derived microglia samples.

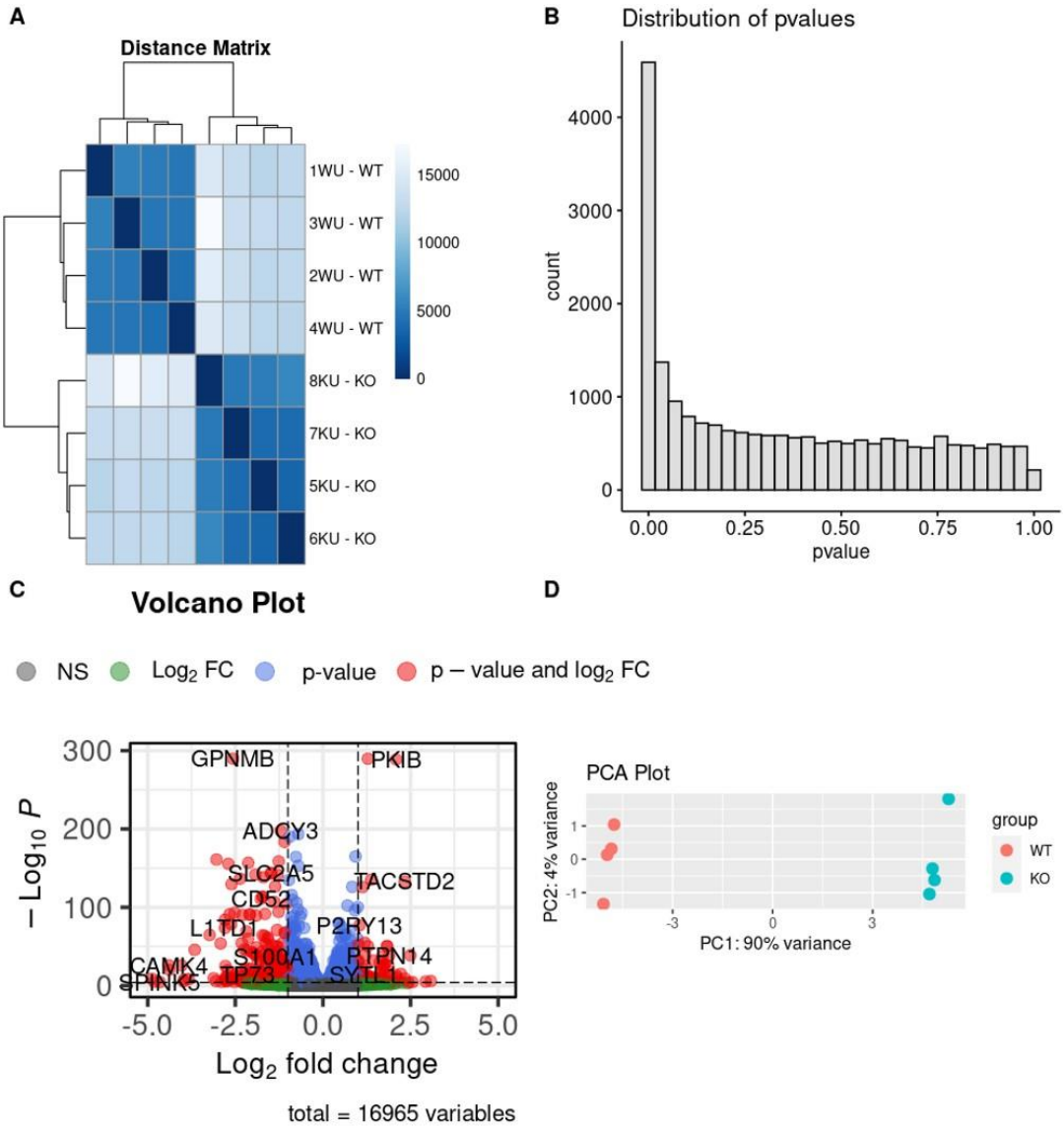

Figure 85 - Summary graphs of PRJNA662330. There are high amount of DEGs and the samples are clearly separated. PC1 has 90% variance and the samples are separated on this axis.

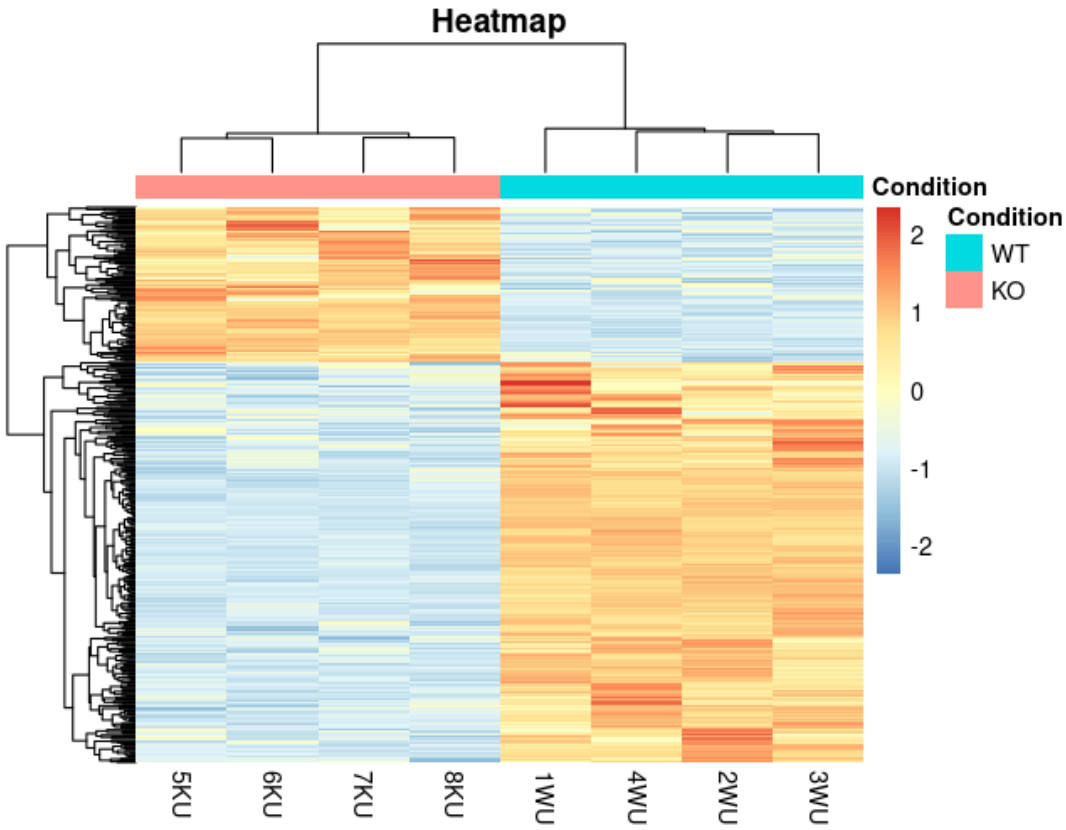

Figure 86 - Heatmap of PRJNA662330. Samples show a very clear expression profile distinction.

8.13. PRJNA675864

This dataset compares healthy and AD blood samples.

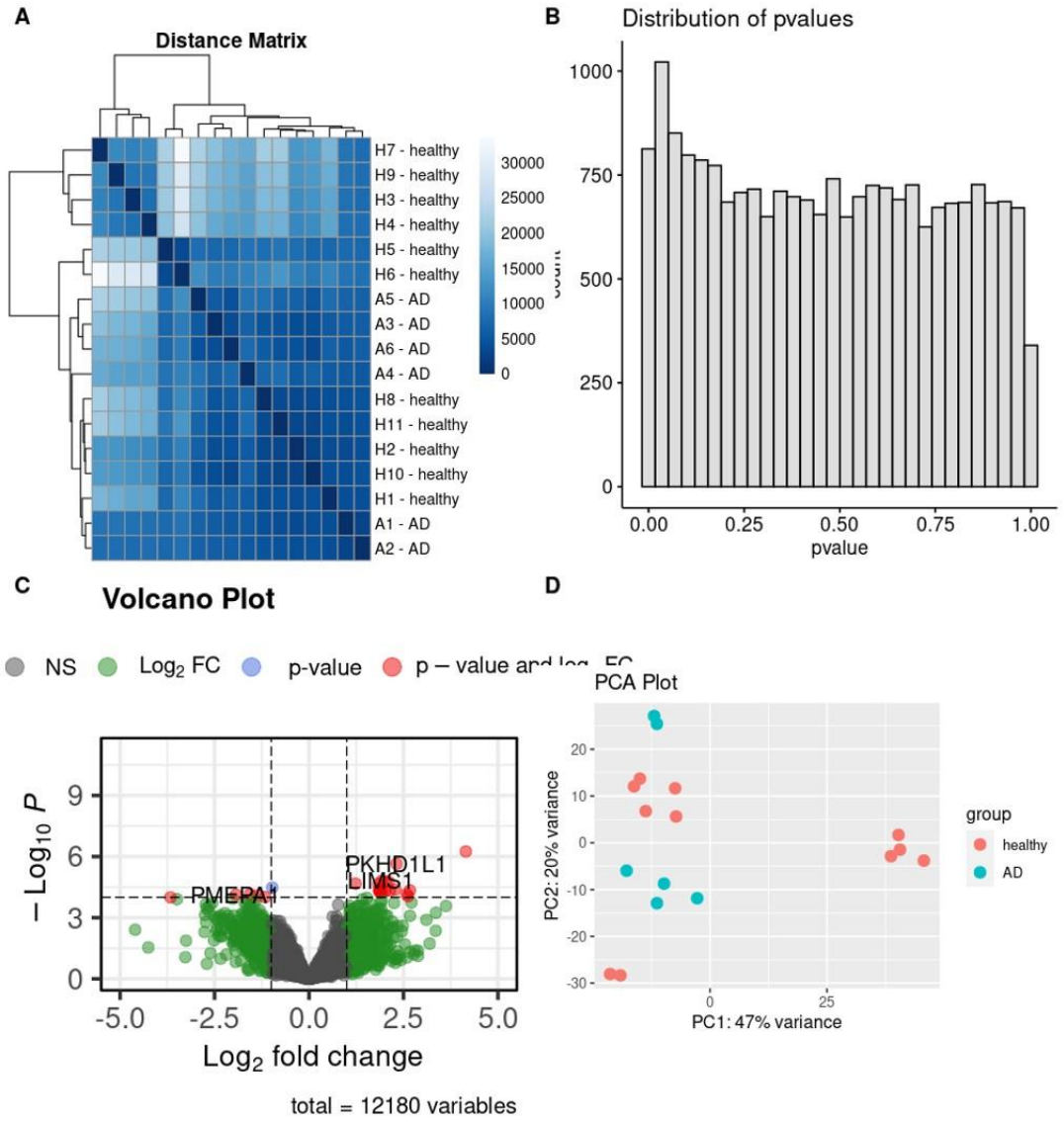

Figure 87 - summary graphs of PRJNA675864. On the PCA plot, most AD samples are positioned on the left but they are mixed with more than half of the healthy samples. Amount of DEGs is low.

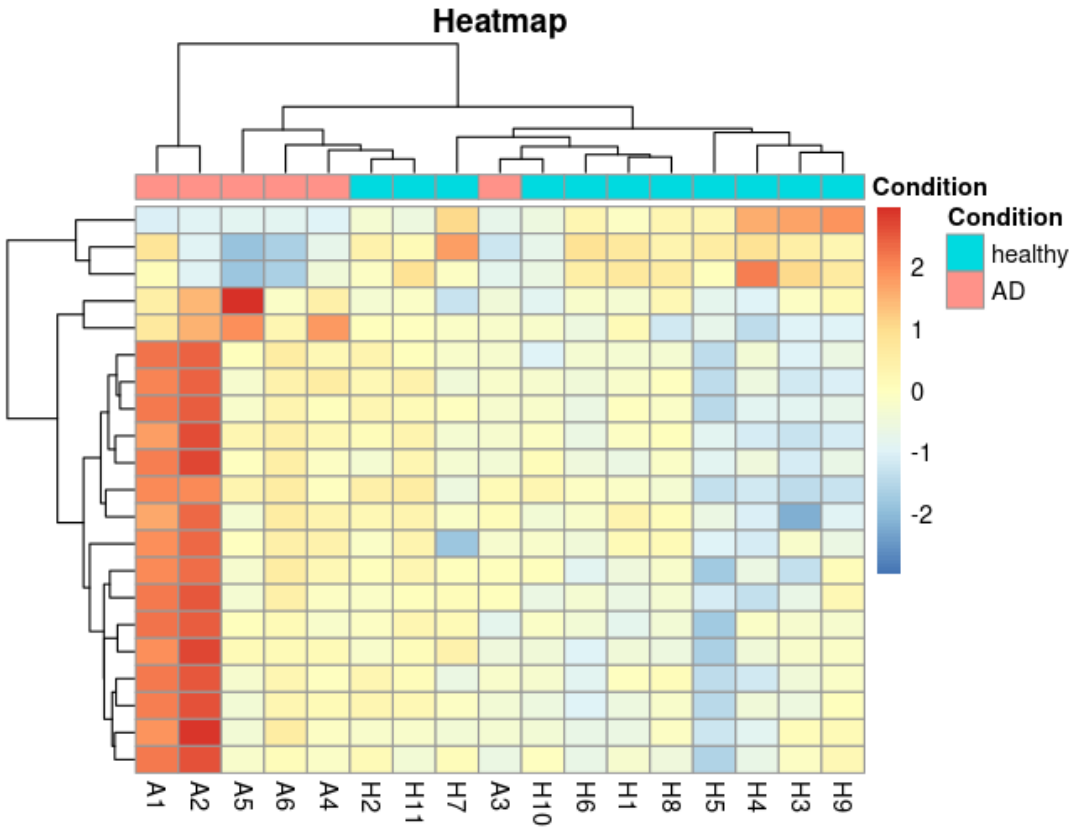

Figure 88 – Heatmap of PRJNA675864. Other than 2, most of the samples have relatively similar expression profile.

8.14. PRJNA683625

This dataset compares healthy and AD-like (iPSCs from AD patients) iPSC-derived neural progenitor cells.

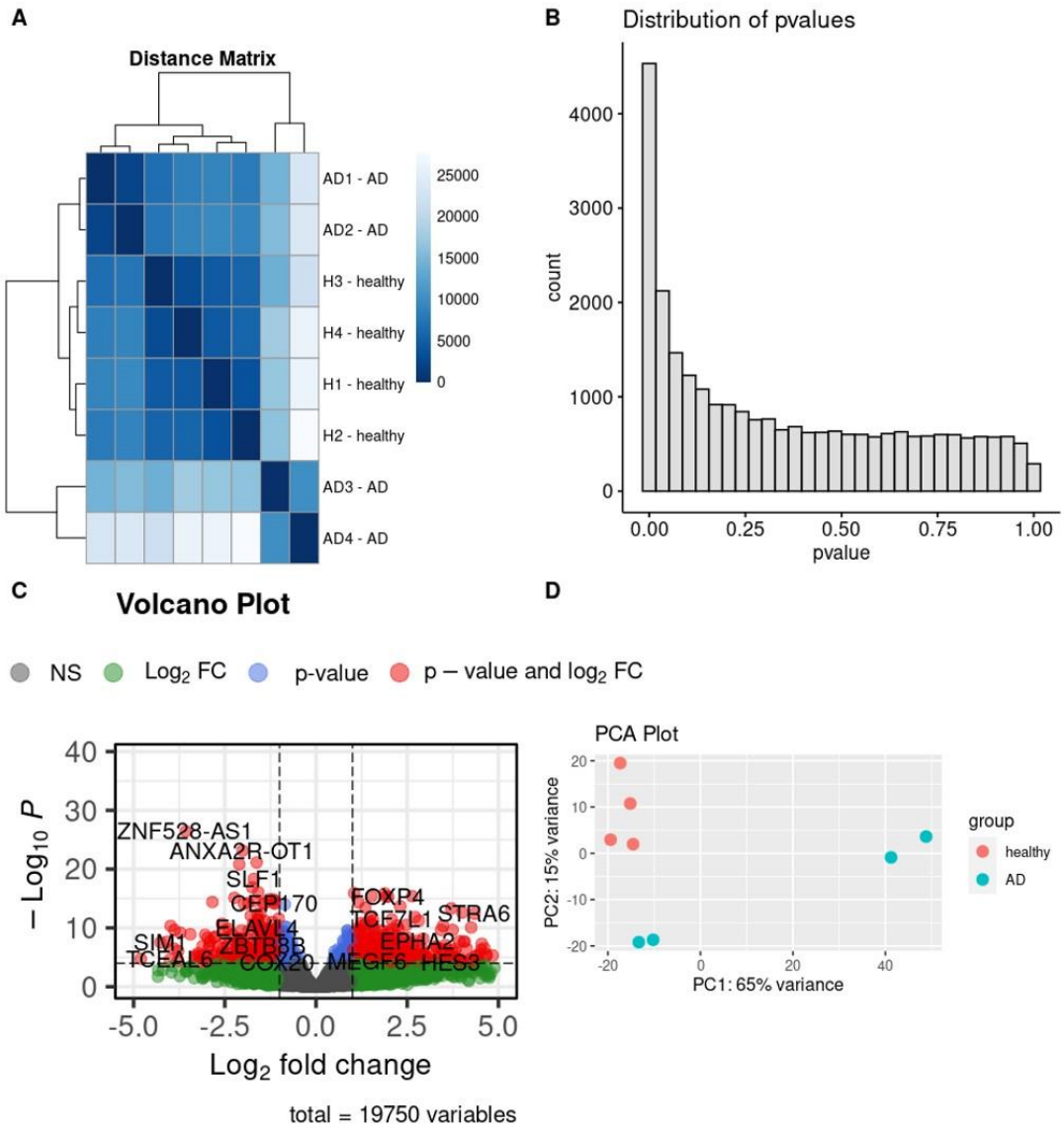

Figure 89 - Summary graphs of PRJNA682625. On both PC1 and PC2, AD samples are mixed with healthy samples. Yet, healthy samples are clustered very close. The amount of DEGs is also high.

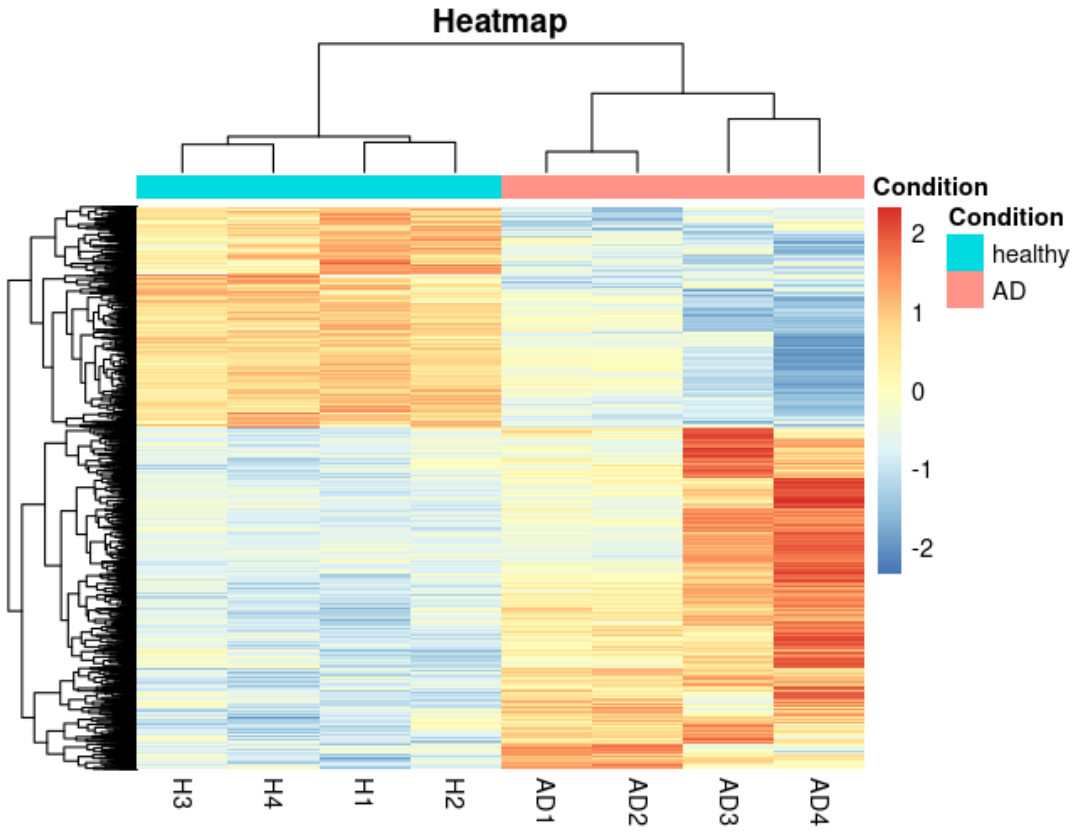

Figure 90 - Heatmap of PRJNA683625. AD and healthy samples show a clear distinction in terms of gene expression profiles.

8.15. PRJNA688060

This dataset compares healthy and AD postmortem middle temporal gyrus samples.

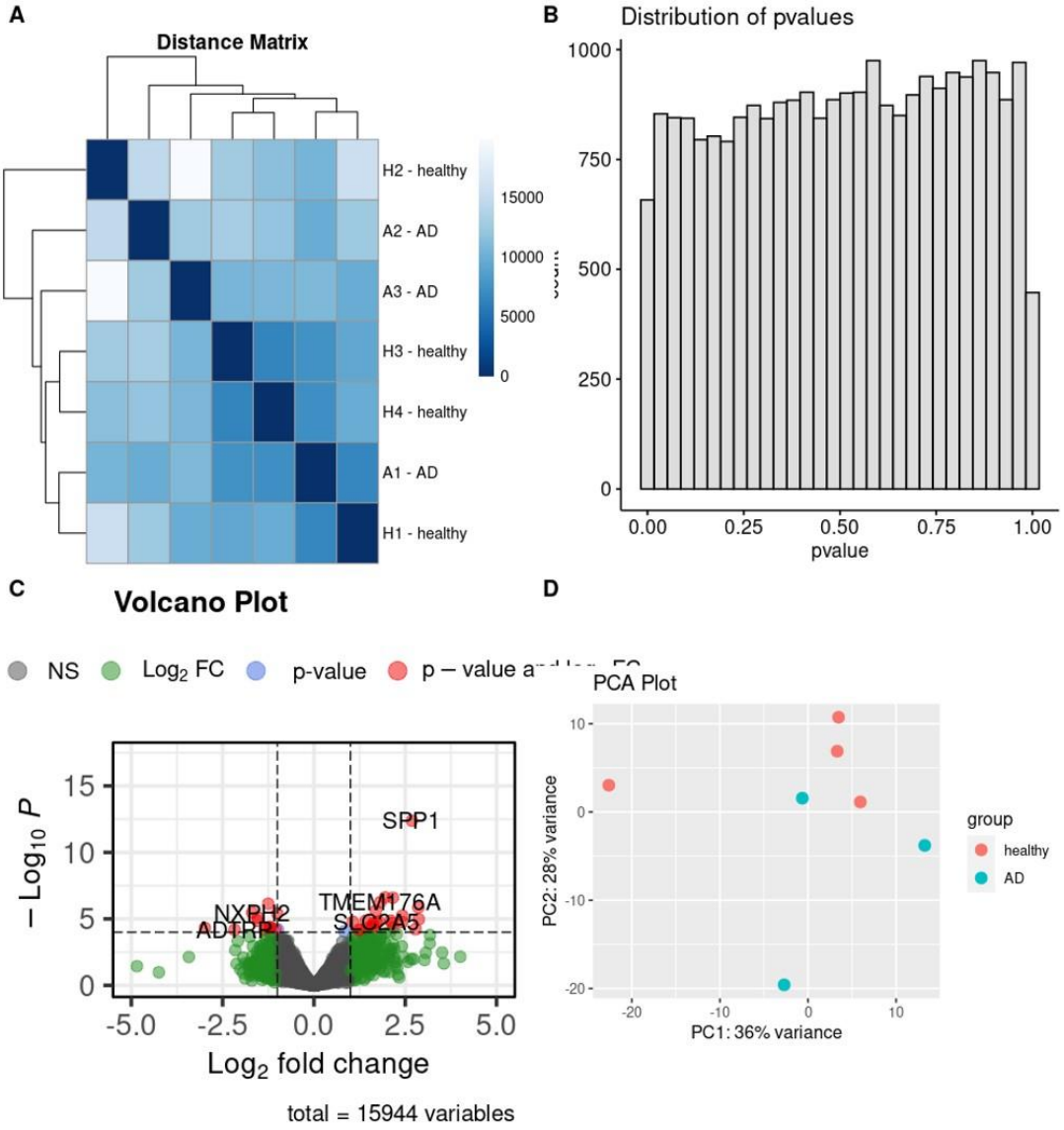

Figure 91 - Summary graphs of PRJNA688060. Healthy and AD samples are mixed together and the amount of DEGs is very low.

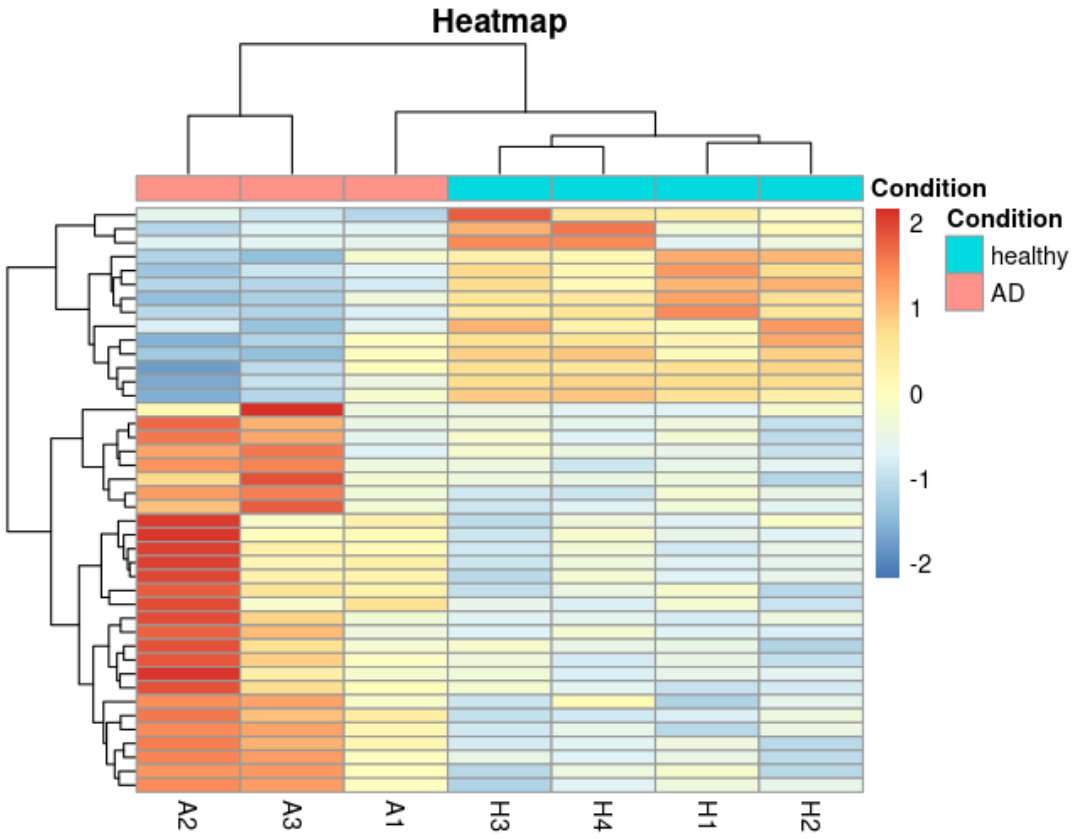

Figure 92 - Heathmap of PRJNA688060. Healthy and AD samples show a distinct expression profile but only for the very low amount of DEGs.

8.16. PRJNA688885

This dataset compares healthy (untreated) and AD-like (serum treated) iPSC-derived cerebral organoids.

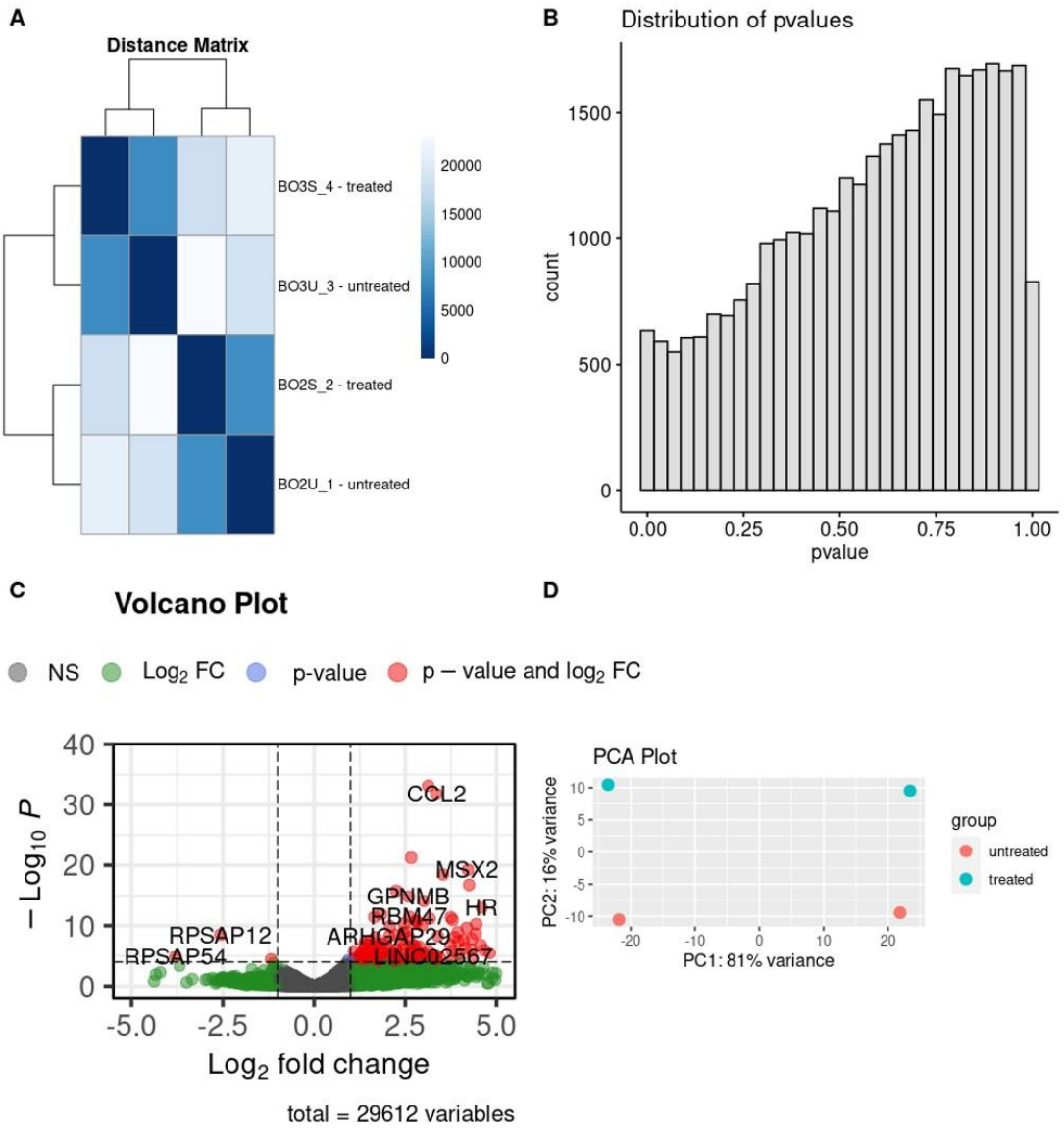

Figure 93 - Summary graphs of PRJNA688885. Samples are separated by PC2 which has only 16% variance. The number of DEGs is low and DEGs are overexpressed.

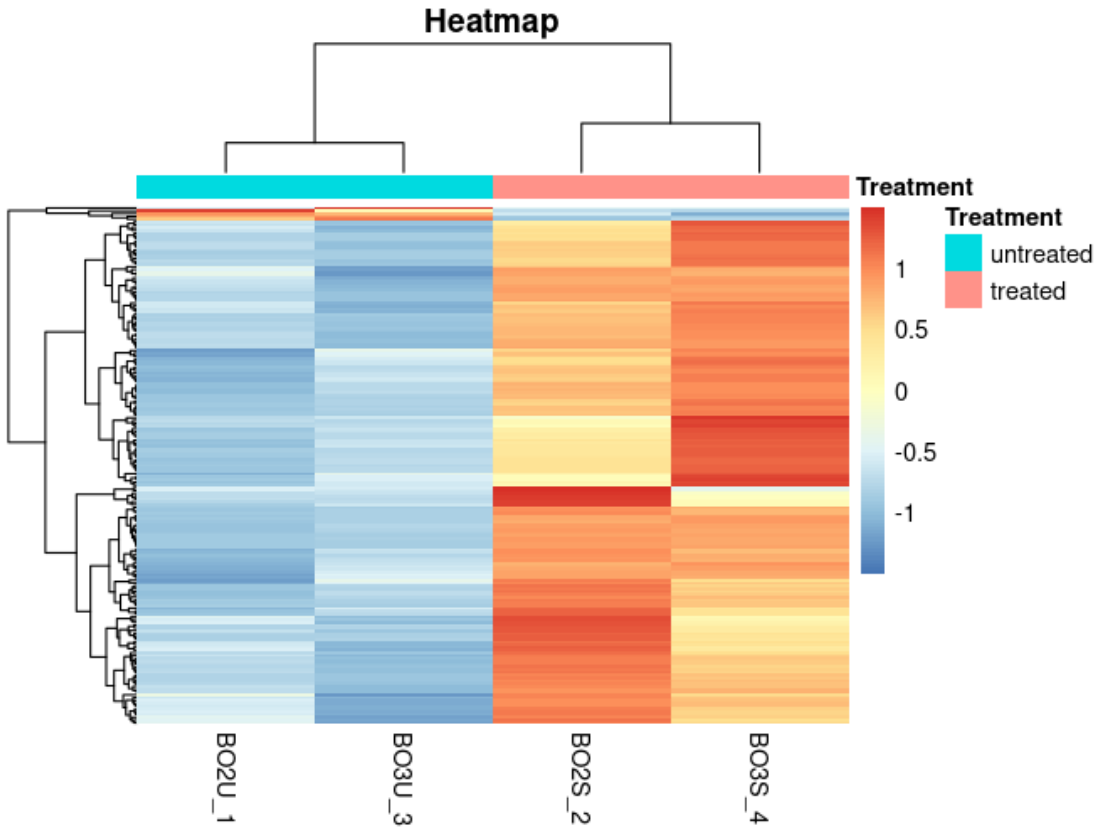

Figure 94 - Heatmap of PRJNA688885. The overexpression profile can be seen in the AD-like cells.

8.17. PRJNA714081

This dataset compares healthy and AD blood samples.

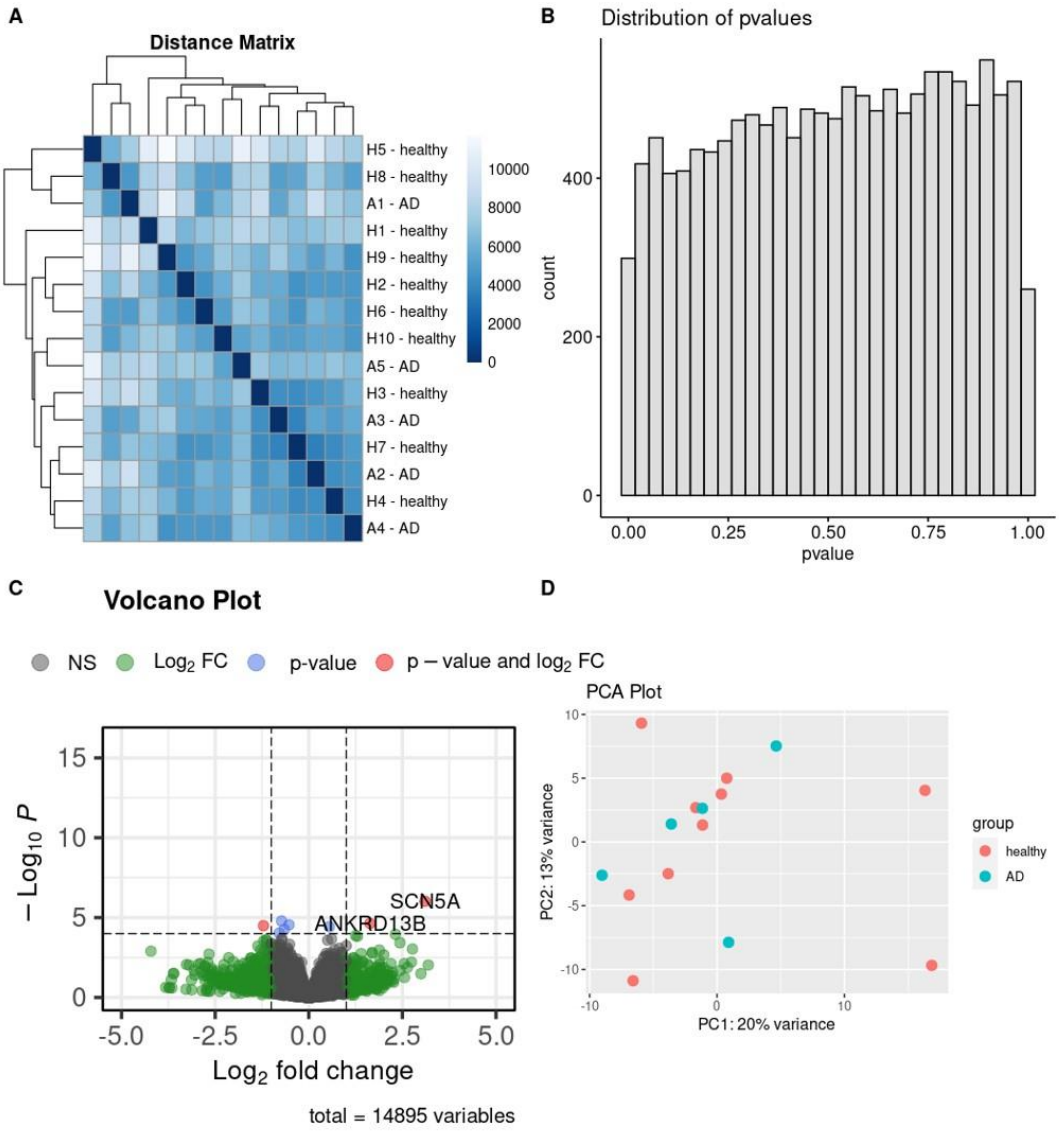

Figure 95 - Summary graphs of PRJNA714081. The amount of DEGs is very low and the samples don't show any distinction. A heatmap cannot be drawn.

8.18. PRJNA727602

This dataset compares healthy and AD postmortem hippocampus samples by considering sexes.

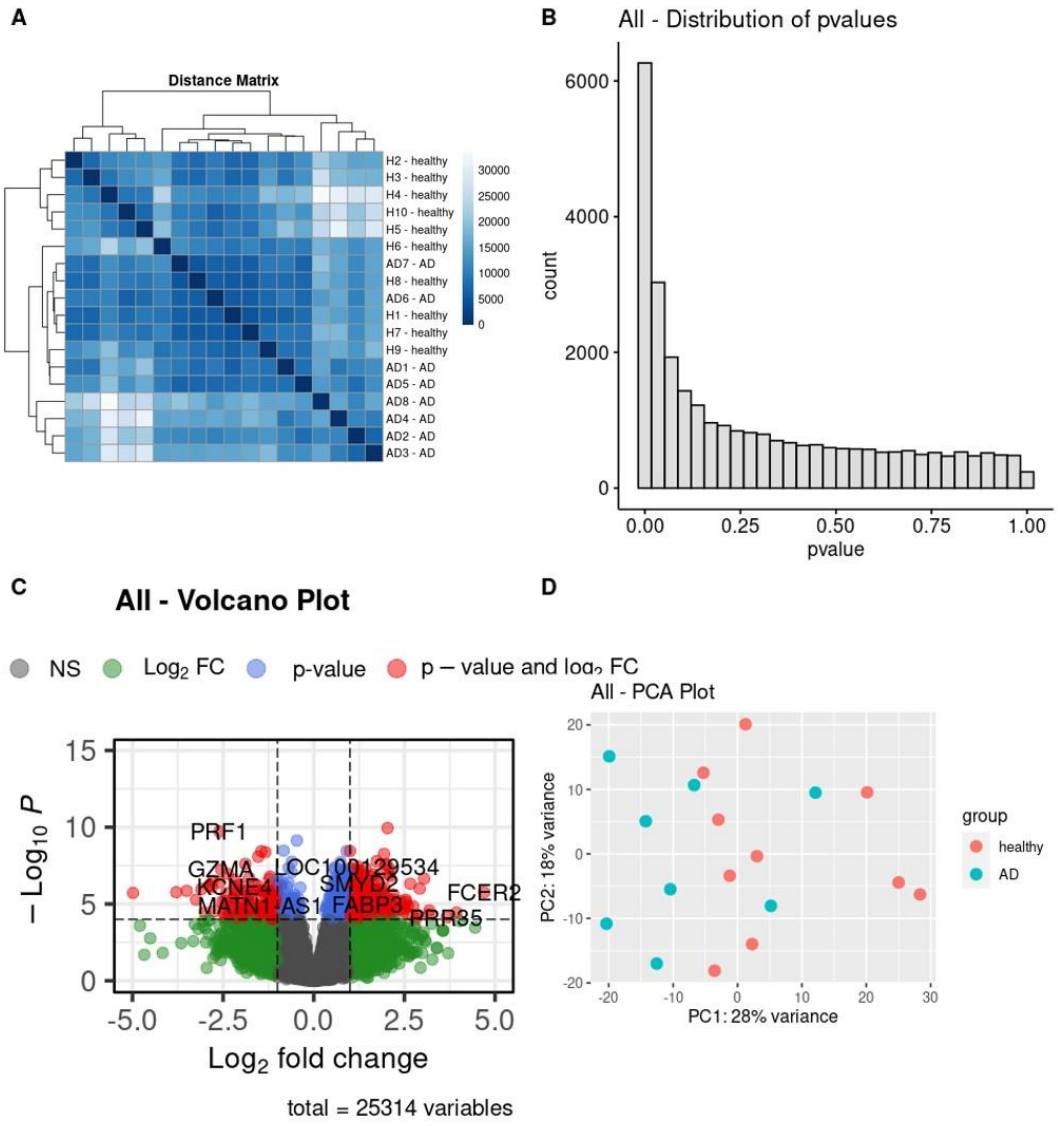

Figure 96 - Summary graphs for PRJNA727602 without considering sex variable. While the AD and healthy cell separation is low, there are high amount of signiciant genes.

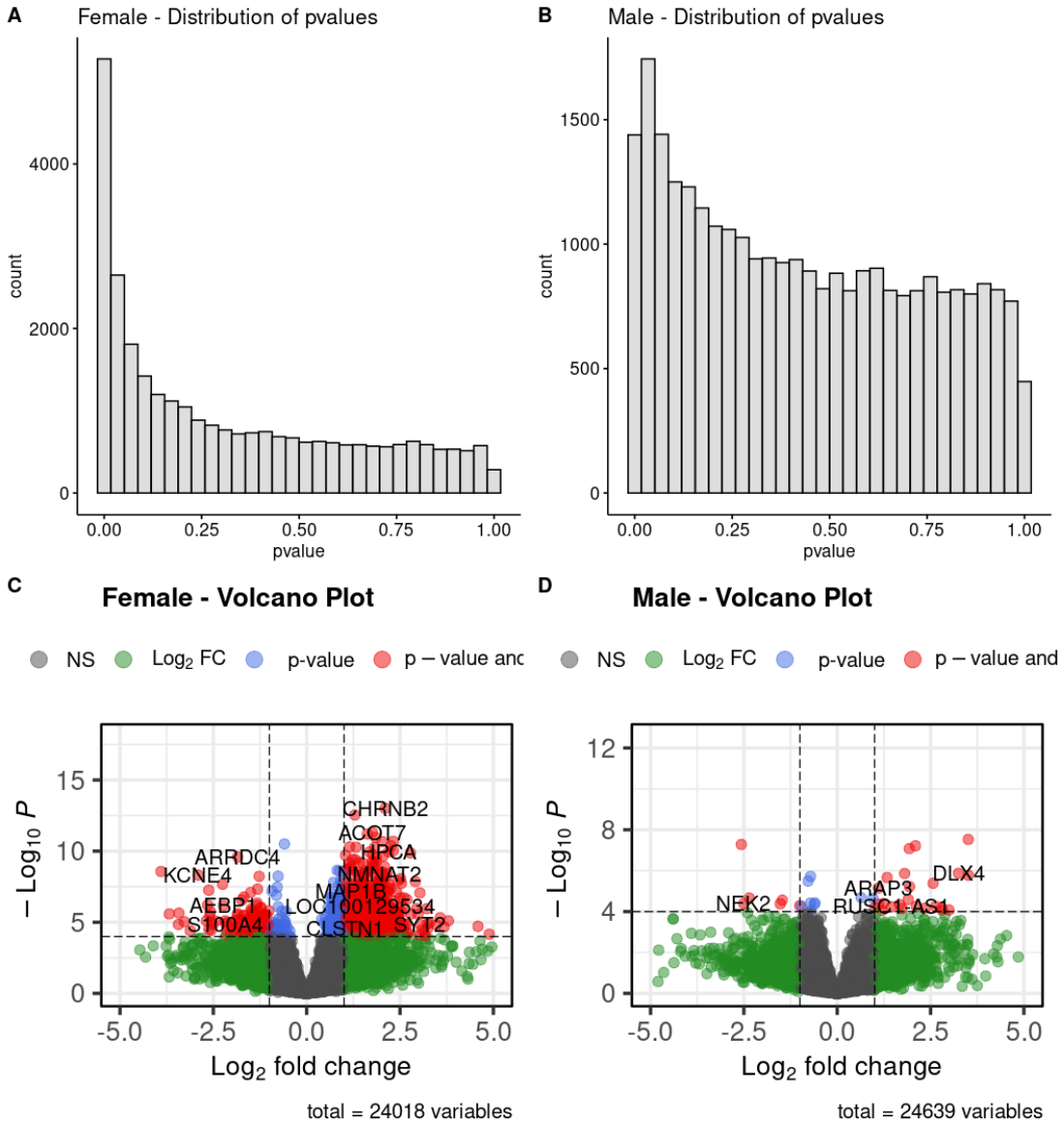

Figure 97 - P-value distribution and volcano plots of female and male samples. Male samples have relatively low amount of significant genes.

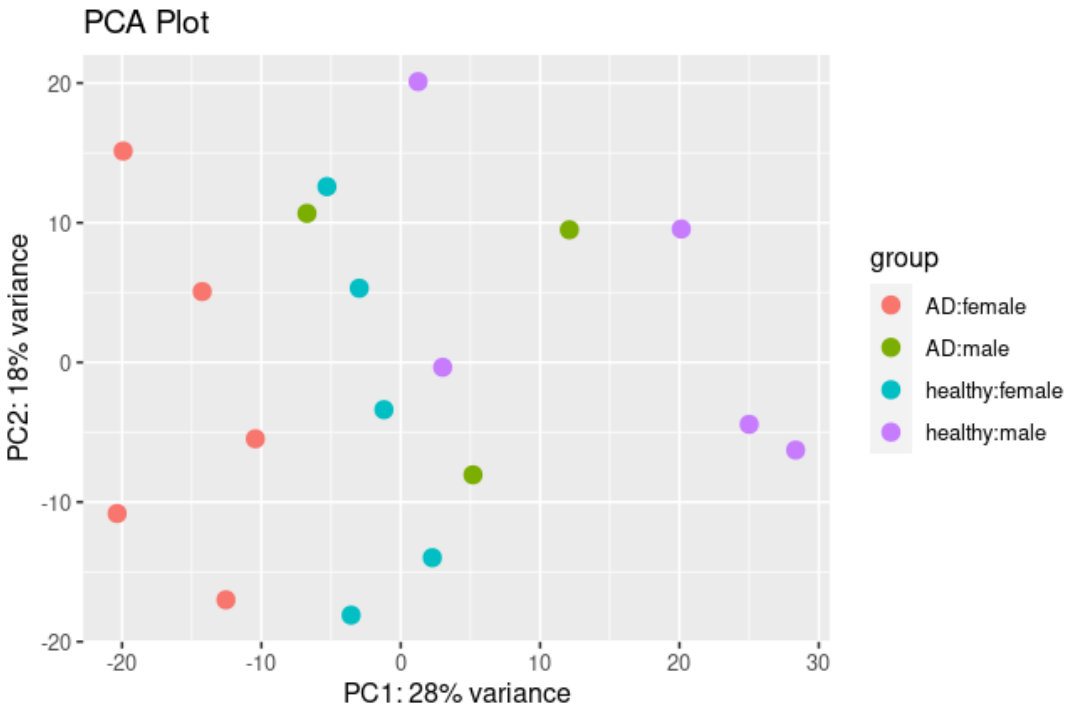

Figure 98 - PCA plot of PRJNA727602. When sexes are added, the distinction is more clear. AD females are most distant to healthy males. The samples are separated by their sex more than their condition.

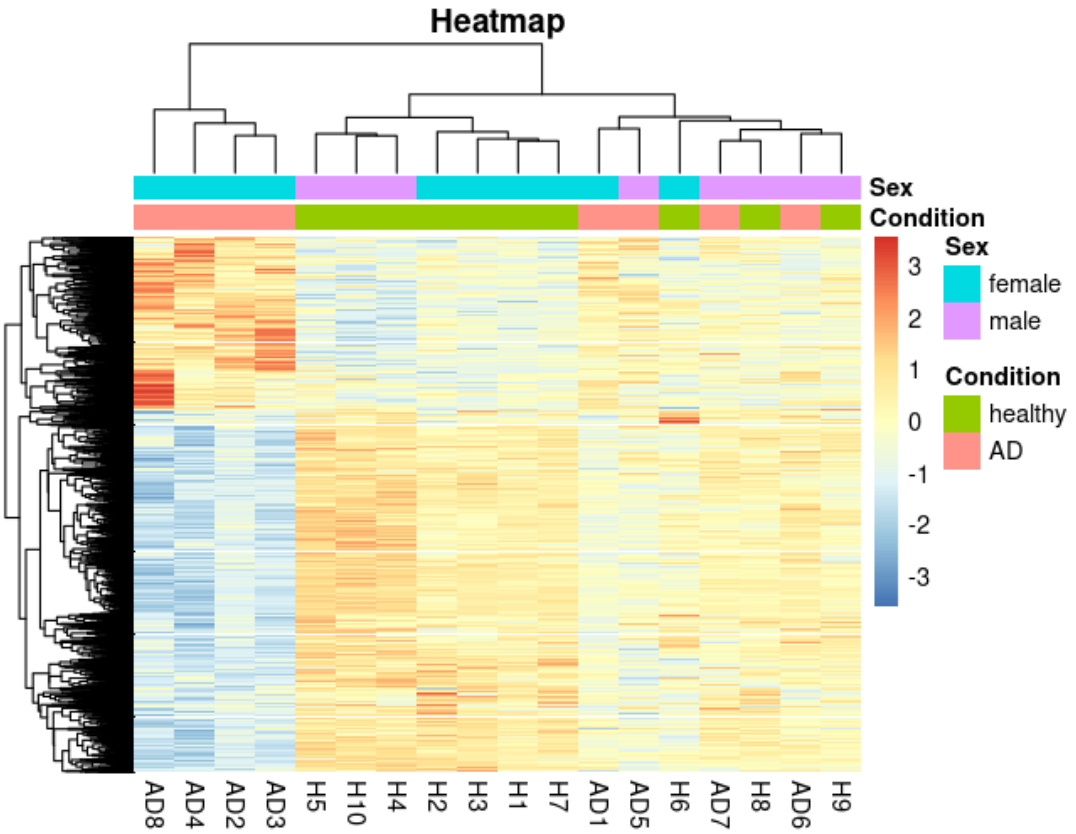

Figure 99 - Heatmap of PRJNA727602. Highest amount of gene expression profile distinction is observed in AD females, like in the PCA plot. Healthy samples are also similar except for 3 samples that are mixed with AD samples.

8.19. PRJNA767074

This dataset compares healthy and AD postmortem hippocampus samples.

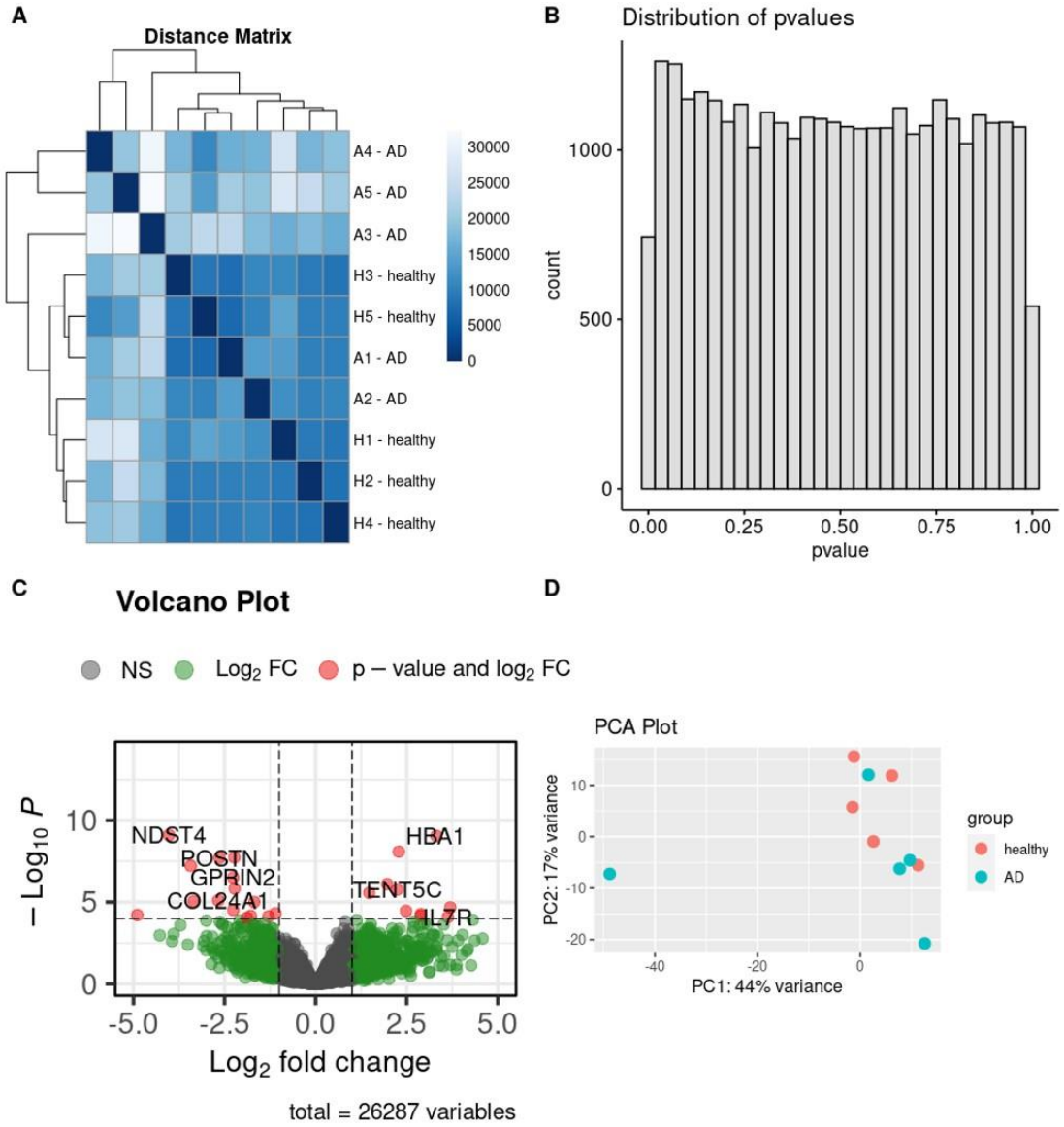

Figure 100 - Summary graphs of PRJNA767074. There are not clear distinction between the samples and the number of DEGs is low.

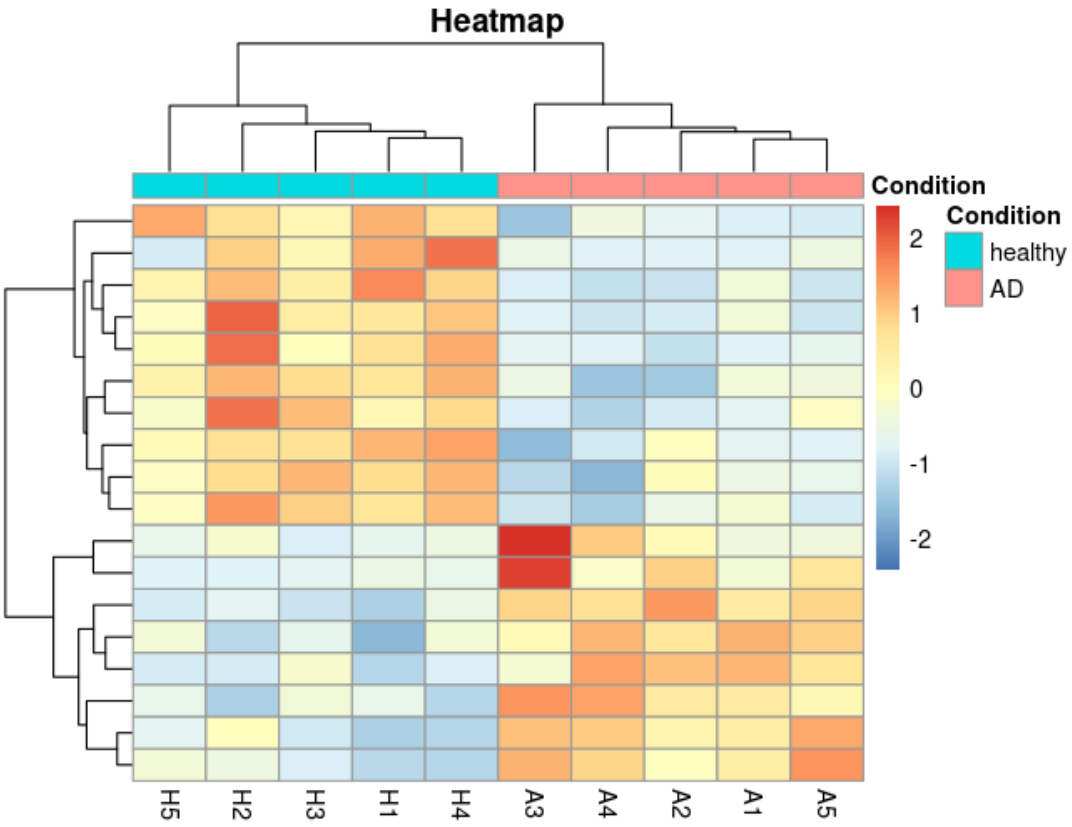

Figure 101 - Heatmap of PRJNA767074. Although this plot shows a clear separation between healthy and AD samples, the number of genes is very low.

APPENDIX C

GSE ANALYSIS RESULTS

9.1. Lists of GO Terms and Tree Plots

9.1.1. Intersecting GO Terms for All Tissues without Sex Variable, not Found in Cell Samples

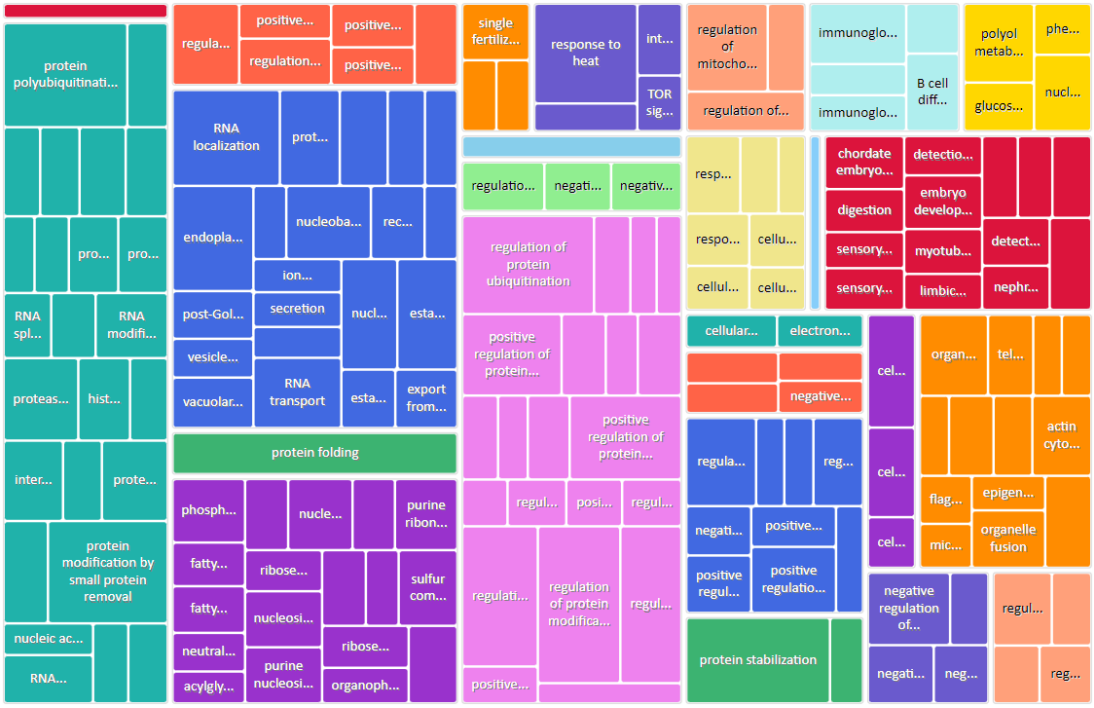

Figure 102 – Enriched BP terms tree plot found in tissues without sex variable and not in cells.

Figure 103 – Enriched CC terms tree plot found in tissues without sex variable and not in cells

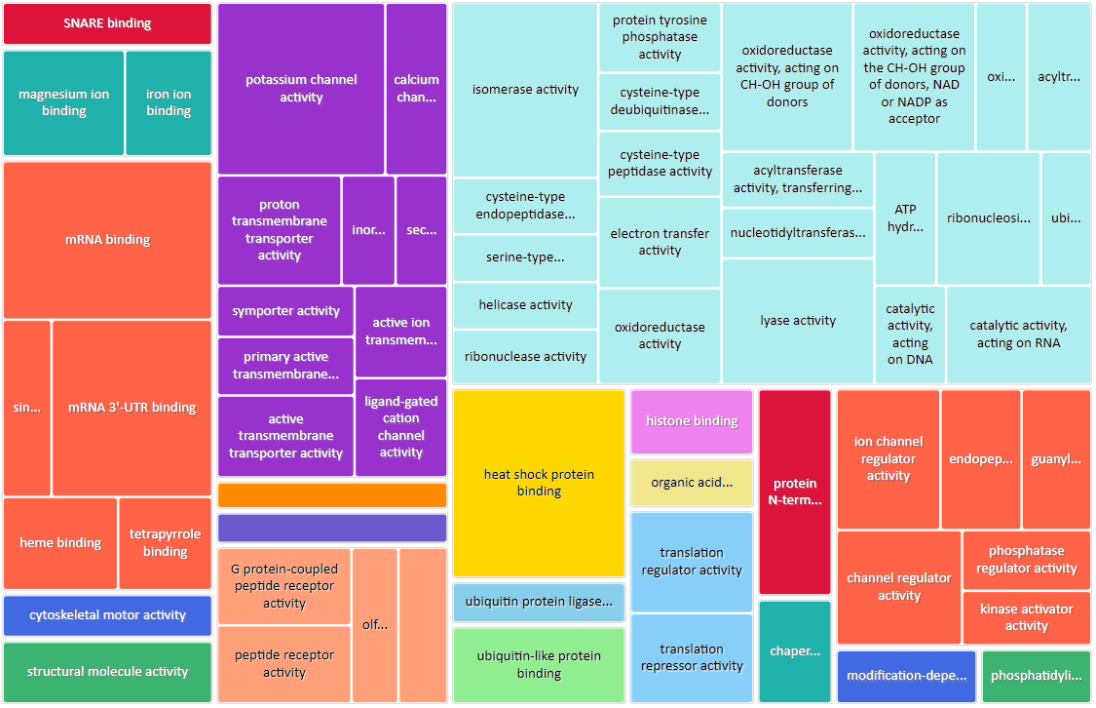

Figure 104 – Enriched MF terms tree plot found in tissues without sex variable and not in cells

Table 29 - Enriched GO terms found in tissues without sex variable and not in cells

TermID	Name	Value	LogSize	Term
GO:0000018	regulation of DNA recombination	-2.758560436	2.127104798	BP
GO:0000045	autophagosome assembly	-3.263283702	1.875061263	BP
GO:0000082	G1/S transition of mitotic cell cycle	-2.450863342	1.934498451	BP
GO:0000209	protein polyubiquitination	-10	2.359835482	BP
GO:0000375	RNA splicing, via transesterification reactions	-2.529737608	2.387389826	BP

GO:0000398	mRNA splicing, via spliceosome	-2.72299291	2.380211242	BP
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	-2.72299291	2.380211242	BP
GO:0000723	telomere maintenance	-2.874837563	1.986771734	BP
GO:0000956	nuclear-transcribed mRNA catabolic process	-3.189629971	2.017033339	BP
GO:0006402	mRNA catabolic process	-2.764300239	2.10720997	BP
GO:0001909	leukocyte mediated cytotoxicity	-2.409638928	1.799340549	BP
GO:0002027	regulation of heart rate	-2.301751454	2.025305865	BP
GO:0002224	toll-like receptor signaling pathway	-2.350698812	1.633468456	BP
GO:0002377	immunoglobulin production	-2.781046768	1.672097858	BP
GO:0002444	myeloid leukocyte mediated immunity	-2.112608815	1.785329835	BP
GO:0002526	acute inflammatory response	-2.289930225	1.892094603	BP
GO:0002702	positive regulation of production of molecular mediator of immune response	-2.597052966	2.133538908	BP
GO:0002703	regulation of leukocyte mediated immunity	-3.443042608	2.380211242	BP
GO:0002705	positive regulation of leukocyte mediated immunity	-2.941547617	2.1430148	BP
GO:0002706	regulation of lymphocyte mediated immunity	-4.373659633	2.247973266	BP
GO:0006006	glucose metabolic process	-2.753418896	2.064457989	BP
GO:0006275	regulation of DNA replication	-1.992771992	2.139879086	BP
GO:0006281	DNA repair	-1.816339184	2.700703717	BP
GO:0006302	double-strand break repair	-1.879878261	2.309630167	BP
GO:0006310	DNA recombination	-1.829330566	2.385606274	BP
GO:0006397	mRNA processing	-2.905525919	2.652246341	BP
GO:0006401	RNA catabolic process	-1.840361144	2.214843848	BP
GO:0006403	RNA localization	-7.958607315	2.26245109	BP
GO:0006457	protein folding	-10	2.346352974	BP
GO:0006476	protein deacetylation	-2.043803675	1.812913357	BP
GO:0006479	protein methylation	-2.942994265	2.176091259	BP
GO:0016571	histone methylation	-2.981664588	1.977723605	BP
GO:0018022	peptidyl-lysine methylation	-2.847557459	1.944482672	BP
GO:0034968	histone lysine methylation	-2.529669106	1.851258349	BP
GO:0006605	protein targeting	-4.441291429	2.394451681	BP
GO:0006612	protein targeting to membrane	-3.551486537	1.963787827	BP
GO:0006626	protein targeting to mitochondrion	-2.762915857	1.785329835	BP
GO:0006631	fatty acid metabolic process	-2.462988889	2.514547753	BP
GO:0006633	fatty acid biosynthetic process	-2.633103873	2.079181246	BP
GO:0006638	neutral lipid metabolic process	-2.179158081	1.991226076	BP
GO:0006639	acylglycerol metabolic process	-1.797798416	1.986771734	BP
GO:0006661	phosphatidylinositol biosynthetic process	-3.586970885	2.117271296	BP
GO:0006839	mitochondrial transport	-2.403991923	2.170261715	BP
GO:0006888	endoplasmic reticulum to Golgi vesicle-mediated transport	-6.567030709	2.130333768	BP
GO:0006892	post-Golgi vesicle-mediated transport	-2.979895289	1.991226076	BP
GO:0006906	vesicle fusion	-2.420918955	2.012837225	BP
GO:0090174	organelle membrane fusion	-2.28106176	2.021189299	BP
GO:0006997	nucleus organization	-2.795653142	2.167317335	BP
GO:0007034	vacuolar transport	-3.147842303	2.212187604	BP
GO:0007041	lysosomal transport	-1.859249114	2.113943352	BP
GO:0007098	centrosome cycle	-2.434199115	1.968482949	BP
GO:0007338	single fertilization	-3.167564297	2.181843588	BP
GO:0007586	digestion	-2.670643806	2.037426498	BP
GO:0007606	sensory perception of chemical stimulus	-2.239077126	2.742725131	BP
GO:0007608	sensory perception of smell	-2.590887098	2.671172843	BP
GO:0050907	detection of chemical stimulus involved in sensory perception	-2.055338025	2.693726949	BP
GO:0050911	detection of chemical stimulus involved in sensory perception of smell	-2.165745595	2.646403726	BP
GO:0008213	protein alkylation	-2.942994265	2.176091259	BP
GO:0008286	insulin receptor signaling pathway	-2.375980201	1.857332496	BP
GO:0008380	RNA splicing	-2.404575683	2.563481085	BP
GO:0008630	intrinsic apoptotic signaling pathway in response to DNA damage	-2.574176541	1.86923172	BP
GO:0009063	amino acid catabolic process	-2.38413523	2.041392685	BP
GO:0009141	nucleoside triphosphate metabolic process	-3.562744999	2.320146286	BP
GO:0009150	purine ribonucleotide metabolic process	-2.347801335	2.587710965	BP
GO:0009259	ribonucleotide metabolic process	-2.376455127	2.608526034	BP
GO:0009152	purine ribonucleotide biosynthetic process	-3.456252947	2.301029996	BP
GO:0006164	purine nucleotide biosynthetic process	-3.193751167	2.371067862	BP
GO:0009260	ribonucleotide biosynthetic process	-2.501904337	2.330413773	BP
GO:0009308	amine metabolic process	-2.243652703	1.982271233	BP

GO:0009408	response to heat	-8.254144805	2.021189299	BP
GO:0009451	RNA modification	-3.641626339	2.250420002	BP
GO:0009566	fertilization	-2.017988467	2.283301229	BP
GO:0009593	detection of chemical stimulus	-2.400954116	2.722633923	BP
GO:0009792	embryo development ending in birth or egg hatching	-3.279641049	2.827369273	BP
GO:0010498	proteasomal protein catabolic process	-4.782516056	2.57634135	BP
GO:0010675	regulation of cellular carbohydrate metabolic process	-1.879776265	2.178976947	BP
GO:0010821	regulation of mitochondrion organization	-5.761953897	2.181843588	BP
GO:0010948	negative regulation of cell cycle process	-2.278253024	2.431363764	BP
GO:0014902	myotube differentiation	-2.744247642	1.880813592	BP
GO:0015931	nucleobase-containing compound transport	-4.785156152	2.342422681	BP
GO:0016064	immunoglobulin mediated immune response	-4.677780705	2.245512668	BP
GO:0019724	B cell mediated immunity	-4.356547324	2.255272505	BP
GO:0016570	histone modification	-3.314919512	2.592176757	BP
GO:0016579	protein deubiquitination	-2.410218156	2.184691431	BP
GO:0017148	negative regulation of translation	-2.244269639	2.222716471	BP
GO:0018393	internal peptidyl-lysine acetylation	-3.707088449	2.075546961	BP
GO:0006473	protein acetylation	-3.430944843	2.158362492	BP
GO:0006475	internal protein amino acid acetylation	-3.869112634	2.08278537	BP
GO:0016573	histone acetylation	-3.259935447	2.045322979	BP
GO:0018394	peptidyl-lysine acetylation	-3.97643675	2.086359831	BP
GO:0018958	phenol-containing compound metabolic process	-2.36540165	1.949390007	BP
GO:0019693	ribose phosphate metabolic process	-2.4625274	2.618048097	BP
GO:0019725	cellular homeostasis	-3.551854626	2.819543936	BP
GO:0019751	polyol metabolic process	-4.474955193	2.021189299	BP
GO:0021761	limbic system development	-2.196862122	2.056904851	BP
GO:0021782	glial cell development	-2.286953808	2.037426498	BP
GO:0022900	electron transport chain	-2.482276148	2.230448921	BP
GO:0030029	actin filament-based process	-4.402304814	2.783188691	BP
GO:0030036	actin cytoskeleton organization	-3.706413076	2.736396502	BP
GO:0030183	B cell differentiation	-3.280716615	2.11058971	BP
GO:0030278	regulation of ossification	-3.368921384	2.086359831	BP
GO:0030317	flagellated sperm motility	-1.956123458	2.06069784	BP
GO:0060294	cilium movement involved in cell motility	-1.943951982	2.093421685	BP
GO:0030518	intracellular steroid hormone receptor signaling pathway	-1.658395083	1.748188027	BP
GO:0031023	microtubule organizing center organization	-1.78616753	2.029383778	BP
GO:0031396	regulation of protein ubiquitination	-10	2.324282455	BP
GO:1903322	positive regulation of protein modification by small protein conjugation or removal	-6.329754147	2.1430148	BP
GO:0031398	positive regulation of protein ubiquitination	-6.254144805	2.08278537	BP
GO:0031623	receptor internalization	-3.033984708	1.86332286	BP
GO:0031929	TOR signaling	-1.995078974	1.579783597	BP
GO:0032411	positive regulation of transporter activity	-1.96062536	2.100370545	BP
GO:0032414	positive regulation of ion transmembrane transporter activity	-2.022483566	2.06069784	BP
GO:0032526	response to retinoic acid	-3.357773257	2.041392685	BP
GO:0033044	regulation of chromosome organization	-2.648246026	2.403120521	BP
GO:0033157	regulation of intracellular protein transport	-3.450348166	2.352182518	BP
GO:0033865	nucleoside bisphosphate metabolic process	-3.358052002	2.089905111	BP
GO:0034032	purine nucleoside bisphosphate metabolic process	-3.358052002	2.089905111	BP
GO:0033875	ribonucleoside bisphosphate metabolic process	-3.358052002	2.089905111	BP
GO:0034219	carbohydrate transmembrane transport	-1.853757722	1.77815125	BP
GO:0034220	ion transmembrane transport	-2.25105319	2.989004616	BP
GO:0034248	regulation of amide metabolic process	-2.781816144	2.701567985	BP
GO:0034249	negative regulation of amide metabolic process	-3.093543952	2.276461804	BP
GO:0034763	negative regulation of transmembrane transport	-2.525800137	2.100370545	BP
GO:0035335	peptidyl-tyrosine dephosphorylation	-2.577012935	1.531478917	BP
GO:0036503	ERAD pathway	-2.425810302	2	BP
GO:0040029	epigenetic regulation of gene expression	-1.989310288	2.113943352	BP
GO:0042176	regulation of protein catabolic process	-2.01628802	2.555094449	BP
GO:0042471	ear morphogenesis	-2.214274038	2.093421685	BP
GO:0042752	regulation of circadian rhythm	-1.984782507	2.086359831	BP
GO:0043009	chordate embryonic development	-3.296146434	2.812913357	BP
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	-4.106793247	2.539076099	BP
GO:0043484	regulation of RNA splicing	-2.745732723	2.257678575	BP
GO:0043487	regulation of RNA stability	-2.321459359	2.252853031	BP

GO:0043488	regulation of mRNA stability	-1.930432881	2.222716471	BP
GO:0043543	protein acylation	-3.523190457	2.264817823	BP
GO:0043549	regulation of kinase activity	-2.716171773	2.894316063	BP
GO:0044242	cellular lipid catabolic process	-1.932988197	2.28780173	BP
GO:0044272	sulfur compound biosynthetic process	-3.458473373	2.195899652	BP
GO:0045089	positive regulation of innate immune response	-2.178585196	2.392696953	BP
GO:0045333	cellular respiration	-2.618091078	2.285557309	BP
GO:0045446	endothelial cell differentiation	-2.499677332	1.934498451	BP
GO:0045732	positive regulation of protein catabolic process	-7.253365801	2.324282455	BP
GO:0045931	positive regulation of mitotic cell cycle	-2.367368012	2.079181246	BP
GO:0046390	ribose phosphate biosynthetic process	-2.827737667	2.344392274	BP
GO:0046434	organophosphate catabolic process	-2.394188187	2.195899652	BP
GO:0046488	phosphatidylinositol metabolic process	-2.935823128	2.198657087	BP
GO:0046683	response to organophosphorus	-2.635448922	2.10720997	BP
GO:0046822	regulation of nucleocytoplasmic transport	-4.850780887	2.037426498	BP
GO:0046887	positive regulation of hormone secretion	-2.93742007	2.123851641	BP
GO:0046903	secretion	-2.357423423	2.720159303	BP
GO:0046916	cellular transition metal ion homeostasis	-2.006035411	2.06069784	BP
GO:0048193	Golgi vesicle transport	-2.09814426	2.46834733	BP
GO:0048284	organelle fusion	-3.310207041	2.10720997	BP
GO:0048640	negative regulation of developmental growth	-2.738974457	2.037426498	BP
GO:0050658	RNA transport	-4.76700389	2.204119983	BP
GO:0050657	nucleic acid transport	-4.76700389	2.204119983	BP
GO:0051028	mRNA transport	-3.028279572	2.113943352	BP
GO:0051236	establishment of RNA localization	-4.815308569	2.209515015	BP
GO:0050671	positive regulation of lymphocyte proliferation	-2.539202499	2.164352856	BP
GO:0032946	positive regulation of mononuclear cell proliferation	-2.33883112	2.173186268	BP
GO:0042129	regulation of T cell proliferation	-2.493956188	2.257678575	BP
GO:0050684	regulation of mRNA processing	-1.636692249	2.149219113	BP
GO:0050821	protein stabilization	-10	2.330413773	BP
GO:0050864	regulation of B cell activation	-2.80918174	2.344392274	BP
GO:0050906	detection of stimulus involved in sensory perception	-2.535204027	2.749736316	BP
GO:0051052	regulation of DNA metabolic process	-2.124977697	2.729974286	BP
GO:0051054	positive regulation of DNA metabolic process	-2.019539657	2.482873584	BP
GO:0051168	nuclear export	-4.774690718	2.123851641	BP
GO:0051928	positive regulation of calcium ion transport	-2.723576172	2.11058971	BP
GO:0055082	cellular chemical homeostasis	-4.4710833	2.741939078	BP
GO:0055086	nucleobase-containing small molecule metabolic process	-3.53375932	2.779596491	BP
GO:0061013	regulation of mRNA catabolic process	-2.144387467	2.271841607	BP
GO:0070646	protein modification by small protein removal	-9.512861625	2.23299611	BP
GO:0071466	cellular response to xenobiotic stimulus	-2.222443597	2.26245109	BP
GO:0072009	nephron epithelium development	-2.337084403	2.021189299	BP
GO:0072594	establishment of protein localization to organelle	-5.127261173	2.526339277	BP
GO:0090150	establishment of protein localization to membrane	-2.435105265	2.328379603	BP
GO:0090305	nucleic acid phosphodiester bond hydrolysis	-2.200836855	2.437750563	BP
GO:0090316	positive regulation of intracellular protein transport	-4.413412695	2.193124598	BP
GO:0090501	RNA phosphodiester bond hydrolysis	-3.351767292	2.235528447	BP
GO:0097237	cellular response to toxic substance	-2.405417248	2.06069784	BP
GO:0097435	supramolecular fiber organization	-3.333555993	2.758911892	BP
GO:0097722	sperm motility	-1.956123458	2.06069784	BP
GO:0099565	chemical synaptic transmission, postsynaptic	-3.007758968	2.068185862	BP
GO:0140352	export from cell	-2.802044739	2.672097858	BP
GO:1901136	carbohydrate derivative catabolic process	-2.187323326	2.243038049	BP
GO:1901343	negative regulation of vasculature development	-2.884377804	2.012837225	BP
GO:1901987	regulation of cell cycle phase transition	-2.006639378	2.631443769	BP
GO:1901991	negative regulation of mitotic cell cycle phase transition	-2.27629936	2.212187604	BP
GO:1901988	negative regulation of cell cycle phase transition	-1.997375978	2.346352974	BP
GO:1902115	regulation of organelle assembly	-3.78267805	2.369215857	BP
GO:1903008	organelle disassembly	-4.330683119	2.017033339	BP
GO:1903038	negative regulation of leukocyte cell-cell adhesion	-4.815308569	2.136720567	BP
GO:0050868	negative regulation of T cell activation	-1.894570223	2.100370545	BP
GO:1903050	regulation of proteolysis involved in protein catabolic process	-8.189767482	2.357934847	BP
GO:1903313	positive regulation of mRNA metabolic process	-2.151105592	2.100370545	BP
GO:1903320	regulation of protein modification by small protein conjugation or removal	-10	2.403120521	BP
GO:1905037	autophagosome organization	-2.444113268	1.913813852	BP

GO:1905477	positive regulation of protein localization to membrane	-2.276102847	2.008600172	BP
GO:1990748	cellular detoxification	-1.906366807	2.025305865	BP
GO:2000058	regulation of ubiquitin-dependent protein catabolic process	-7.40560745	2.240549248	BP
GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	-5.88941029	2.155336037	BP
GO:0061136	regulation of proteasomal protein catabolic process	-6.863279433	2.285557309	BP
GO:1901800	positive regulation of proteasomal protein catabolic process	-4.17327748	2.075546961	BP
GO:1903052	positive regulation of proteolysis involved in protein catabolic process	-5.490797478	2.1430148	BP
GO:2000060	positive regulation of ubiquitin-dependent protein catabolic process	-4.08249449	2.056904851	BP
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	-2.570272111	2.269512944	BP
GO:2000278	regulation of DNA biosynthetic process	-2.240324241	2.100370545	BP
GO:2001020	regulation of response to DNA damage stimulus	-2.442658559	2.496929648	BP
GO:2001242	regulation of intrinsic apoptotic signaling pathway	-3.474773811	2.220108088	BP
GO:0000151	ubiquitin ligase complex	-2.008622175	2.498310554	CC
GO:0000228	nuclear chromosome	-2.34110347	2.367355921	CC
GO:0000781	chromosome, telomeric region	-2.435508201	2.187520721	CC
GO:0000922	spindle pole	-1.775635608	2.243038049	CC
GO:0005681	spliceosomal complex	-3.620228611	2.29666519	CC
GO:0005741	mitochondrial outer membrane	-3.357738498	2.320146286	CC
GO:0031968	organelle outer membrane	-2.845406269	2.374748346	CC
GO:0005759	mitochondrial matrix	-3.990150669	2.685741739	CC
GO:0005776	autophagosome	-1.927955475	2.075546961	CC
GO:0005777	peroxisome	-4.958607315	2.167317335	CC
GO:0005813	centrosome	-2.901779963	2.805500858	CC
GO:0005814	centriole	-2.926101747	2.209515015	CC
GO:0005929	cilium	-3.336483508	2.867467488	CC
GO:0016605	PML body	-2.031305812	2.025305865	CC
GO:0016607	nuclear speck	-2.727038108	2.621176282	CC
GO:0019866	organelle inner membrane	-1.871241081	2.749736316	CC
GO:0019867	outer membrane	-2.911838707	2.378397901	CC
GO:0030175	filopodium	-3.238506316	2.037426498	CC
GO:0030665	clathrin-coated vesicle membrane	-4.612610174	2.130333768	CC
GO:0030672	synaptic vesicle membrane	-2.213122348	2.08278537	CC
GO:0030863	cortical cytoskeleton	-1.95574512	2.037426498	CC
GO:0030880	RNA polymerase complex	-2.334980001	2.093421685	CC
GO:0031461	cullin-RING ubiquitin ligase complex	-7.053547735	2.269512944	CC
GO:0031965	nuclear membrane	-6.534617149	2.487138375	CC
GO:0034399	nuclear periphery	-1.505043015	2.170261715	CC
GO:0034451	centriolar satellite	-2.770576105	2.064457989	CC
GO:0042579	microbody	-4.958607315	2.167317335	CC
GO:0044853	plasma membrane raft	-1.957202595	2.06069784	CC
GO:0045111	intermediate filament cytoskeleton	-2.307345027	2.414973348	CC
GO:0072562	blood microparticle	-3.615363376	2.161368002	CC
GO:0097730	non-motile cilium	-1.917631218	2.227886705	CC
GO:0098562	cytoplasmic side of membrane	-1.638143384	2.315970345	CC
GO:0099081	supramolecular polymer	-5.400116928	3.020775488	CC
GO:0099501	exocytic vesicle membrane	-2.213122348	2.08278537	CC
GO:0099513	polymeric cytoskeletal fiber	-2.778483939	2.902546779	CC
GO:0099512	supramolecular fiber	-5.216811309	3.017033339	CC
GO:1902911	protein kinase complex	-2.770456691	2.176091259	CC
GO:1905368	peptidase complex	-5.89279003	2.10720997	CC
GO:1990204	oxidoreductase complex	-2.861421534	2.120573931	CC
GO:0000149	SNARE binding	-2.956122055	2.049218023	MF
GO:0000287	magnesium ion binding	-4.040005162	2.354108439	MF
GO:0001653	peptide receptor activity	-3.206328674	2.187520721	MF
GO:0003697	single-stranded DNA binding	-2.609253569	2.103803721	MF
GO:0003729	mRNA binding	-10	2.509202522	MF
GO:0003730	mRNA 3'-UTR binding	-8.571865206	2.008600172	MF
GO:0003774	cytoskeletal motor activity	-2.91468855	2.06069784	MF
GO:0004197	cysteine-type endopeptidase activity	-2.434844225	2.089905111	MF
GO:0004252	serine-type endopeptidase activity	-2.067714964	2.250420002	MF
GO:0008236	serine-type peptidase activity	-2.057803901	2.292256071	MF
GO:0004386	helicase activity	-2.048473999	2.195899652	MF
GO:0004540	ribonuclease activity	-2.371652114	2.136720567	MF
GO:0004725	protein tyrosine phosphatase activity	-2.567064448	2.004321374	MF
GO:0004843	cysteine-type deubiquitinase activity	-2.126124734	2.056904851	MF

GO:0101005	deubiquitinase activity	-2.040434489	2.086359831	MF
GO:0004866	endopeptidase inhibitor activity	-3.454424765	2.26245109	MF
GO:0004984	olfactory receptor activity	-2.231994197	2.644438589	MF
GO:0005085	guanyl-nucleotide exchange factor activity	-2.964307655	2.36361198	MF
GO:0005198	structural molecule activity	-4.146301788	2.893206753	MF
GO:0005262	calcium channel activity	-3.233467875	2.071882007	MF
GO:0005267	potassium channel activity	-8.709965389	2.096910013	MF
GO:0005506	iron ion binding	-2.886356454	2.187520721	MF
GO:0008134	transcription factor binding	-2.124218178	2.766412847	MF
GO:0008234	cysteine-type peptidase activity	-2.40357195	2.283301229	MF
GO:0008289	lipid binding	-2.354039058	2.925827575	MF
GO:0008528	G protein-coupled peptide receptor activity	-3.198146945	2.170261715	MF
GO:0009055	electron transfer activity	-3.349621886	2.075546961	MF
GO:0015078	proton transmembrane transporter activity	-4.11463878	2.152288344	MF
GO:0015103	inorganic anion transmembrane transporter activity	-1.786803032	2.206825876	MF
GO:0015291	secondary active transmembrane transporter activity	-1.723152465	2.409933123	MF
GO:0015293	symporter activity	-2.106499386	2.164352856	MF
GO:0015399	primary active transmembrane transporter activity	-2.415277154	2.220108088	MF
GO:0016247	channel regulator activity	-4.363512104	2.164352856	MF
GO:0016491	oxidoreductase activity	-3.507679883	2.867467488	MF
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	-5.823908741	2.1430148	MF
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	-5.415668776	2.11058971	MF
GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	-2.270621388	2.264817823	MF
GO:0016746	acyltransferase activity	-2.851508881	2.409933123	MF
GO:0016747	acyltransferase activity, transferring groups other than amino-acyl groups	-2.544253527	2.361727836	MF
GO:0016779	nucleotidyltransferase activity	-2.291881082	2.187520721	MF
GO:0016829	lyase activity	-5.714442691	2.309630167	MF
GO:0016853	isomerase activity	-7.575118363	2.230448921	MF
GO:0016887	ATP hydrolysis activity	-2.46322012	2.551449998	MF
GO:0017111	ribonucleoside triphosphate phosphatase activity	-4.145086978	2.836956737	MF
GO:0016462	pyrophosphatase activity	-4.534617149	2.875639937	MF
GO:0019208	phosphatase regulator activity	-2.368524733	2.06069784	MF
GO:0019209	kinase activator activity	-2.124036126	2.127104798	MF
GO:0019783	ubiquitin-like protein peptidase activity	-1.998921618	2.139879086	MF
GO:0020037	heme binding	-3.234187835	2.149219113	MF
GO:0022804	active transmembrane transporter activity	-3.375797962	2.62838893	MF
GO:0022853	active ion transmembrane transporter activity	-2.582698635	2.456366033	MF
GO:0030371	translation repressor activity	-3.437530698	1.431363764	MF
GO:0030594	neurotransmitter receptor activity	-2.346374443	2.049218023	MF
GO:0031072	heat shock protein binding	-10	2.096910013	MF
GO:0031625	ubiquitin protein ligase binding	-2.309544207	2.477121255	MF
GO:0032182	ubiquitin-like protein binding	-4.175874166	2.08278537	MF
GO:0042393	histone binding	-2.615233435	2.411619706	MF
GO:0043177	organic acid binding	-1.938099532	2.173186268	MF
GO:0045182	translation regulator activity	-3.876483457	2.161368002	MF
GO:0046906	tetrapyrrole binding	-2.637026636	2.178976947	MF
GO:0047485	protein N-terminus binding	-4.798602876	2.029383778	MF
GO:0051087	chaperone binding	-2.456209696	2.025305865	MF
GO:0099094	ligand-gated cation channel activity	-2.771385819	2.075546961	MF
GO:0099106	ion channel regulator activity	-4.419075024	2.149219113	MF
GO:0140030	modification-dependent protein binding	-2.449366454	2.23299611	MF
GO:0140097	catalytic activity, acting on DNA	-2.101000513	2.399673721	MF
GO:0140098	catalytic activity, acting on RNA	-4.255707017	2.600972896	MF
GO:1901981	phosphatidylinositol phosphate binding	-1.935894949	2.260071388	MF

9.1.2. Intersecting GO Terms for All Female Tissues, not Found in Cell Samples

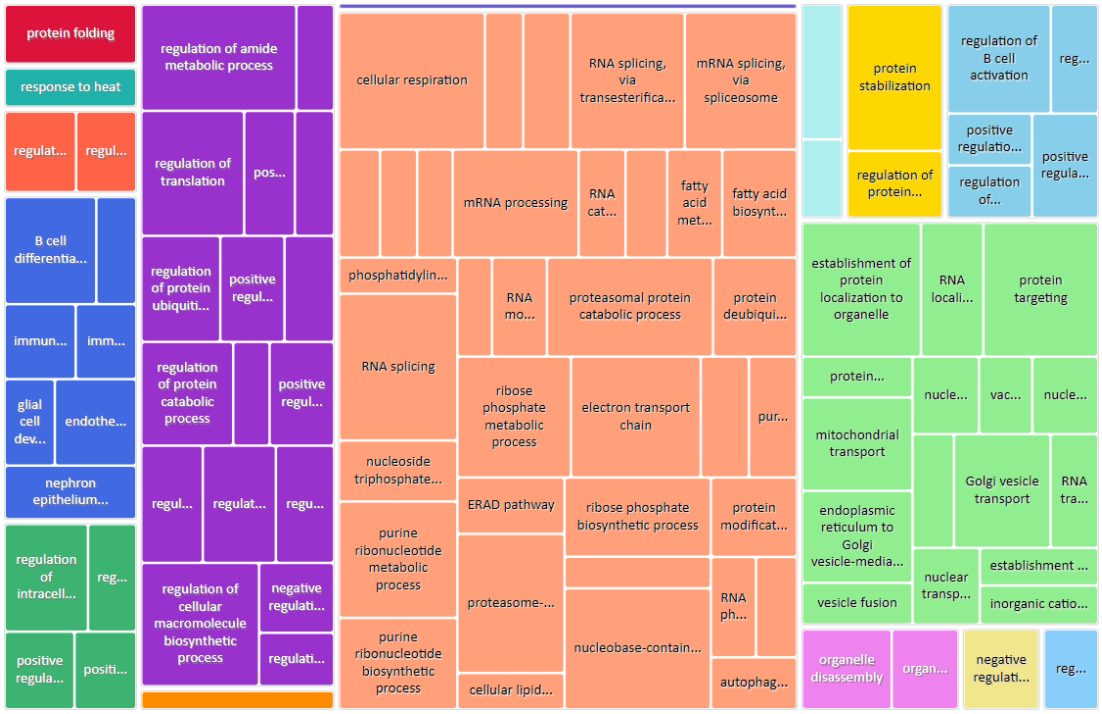

Figure 105 – Enriched BP terms tree plot foound in female tissues and not in cells

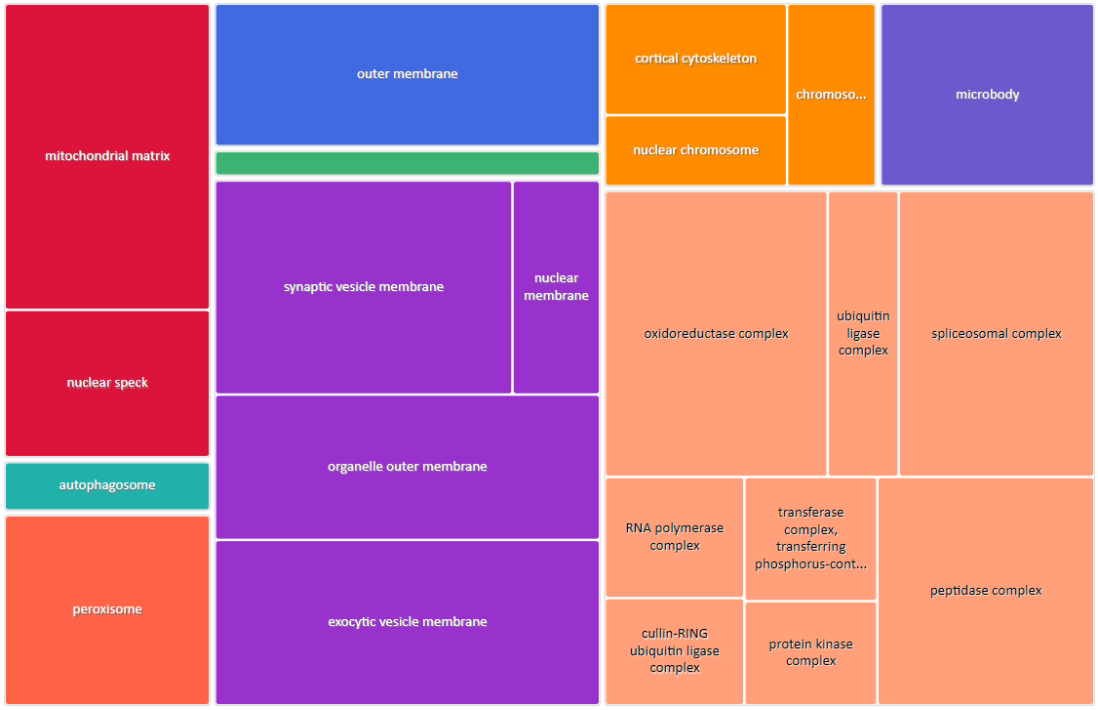

Figure 106 – Enriched CC terms tree plot found in female tissues and not in cells

Figure 107 – Enriched MF terms tree plot found in female tissues and not in cells

Table 30 - Enriched GO terms found in female tissues and not in cells

TermID	Name	Value	LogSize	Term
GO:0000018	regulation of DNA recombination	-2.030672141	2.127104798	BP
GO:0000045	autophagosome assembly	-2.612479234	1.875061263	BP
GO:0000209	protein polyubiquitination	-3.23373812	2.359835482	BP
GO:0000375	RNA splicing, via transesterification reactions	-7.838631998	2.387389826	BP
GO:0000398	mRNA splicing, via spliceosome	-7.684029655	2.380211242	BP
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	-7.684029655	2.380211242	BP
GO:0000723	telomere maintenance	-2.219035178	1.986771734	BP
GO:0001909	leukocyte mediated cytotoxicity	-2.219674719	1.799340549	BP
GO:0002377	immunoglobulin production	-2.755101266	1.672097858	BP
GO:0002703	regulation of leukocyte mediated immunity	-2.677754221	2.380211242	BP
GO:0002705	positive regulation of leukocyte mediated immunity	-2.275854899	2.1430148	BP
GO:0002706	regulation of lymphocyte mediated immunity	-2.303065875	2.247973266	BP
GO:0006006	glucose metabolic process	-2.006084529	2.064457989	BP
GO:0006310	DNA recombination	-1.917863759	2.385606274	BP
GO:0006397	mRNA processing	-6.809668302	2.652246341	BP
GO:0006401	RNA catabolic process	-2.535551418	2.214843848	BP
GO:0006402	mRNA catabolic process	-2.251462414	2.10720997	BP
GO:0006403	RNA localization	-4.226213555	2.26245109	BP
GO:0006417	regulation of translation	-6.528708289	2.64738297	BP
GO:0006457	protein folding	-4.280668713	2.346352974	BP
GO:0006605	protein targeting	-7.782516056	2.394451681	BP
GO:0006612	protein targeting to membrane	-2.284793874	1.963787827	BP
GO:0006631	fatty acid metabolic process	-2.975257111	2.514547753	BP
GO:0006633	fatty acid biosynthetic process	-4.04431225	2.079181246	BP
GO:0006661	phosphatidylinositol biosynthetic process	-2.138312927	2.117271296	BP
GO:0006839	mitochondrial transport	-5.218244625	2.170261715	BP
GO:0006888	endoplasmic reticulum to Golgi vesicle-mediated transport	-5.146301788	2.130333768	BP
GO:0006906	vesicle fusion	-2.334183971	2.012837225	BP
GO:0090174	organelle membrane fusion	-2.118229923	2.021189299	BP
GO:0006913	nucleocytoplasmic transport	-2.603359112	2.390935107	BP
GO:0051168	nuclear export	-2.520184883	2.123851641	BP
GO:0007034	vacuolar transport	-2.086195465	2.212187604	BP

GO:0008380	RNA splicing	-8.68613278	2.563481085	BP
GO:0009141	nucleoside triphosphate metabolic process	-3.613099475	2.320146286	BP
GO:0009150	purine ribonucleotide metabolic process	-6.886056648	2.587710965	BP
GO:0009259	ribonucleotide metabolic process	-6.026410377	2.608526034	BP
GO:0009152	purine ribonucleotide biosynthetic process	-5.407823243	2.301029996	BP
GO:0006164	purine nucleotide biosynthetic process	-5.335358024	2.371067862	BP
GO:0009260	ribonucleotide biosynthetic process	-5.571865206	2.330413773	BP
GO:0009308	amine metabolic process	-1.719757793	1.982271233	BP
GO:0009408	response to heat	-2.889872181	2.021189299	BP
GO:0009451	RNA modification	-2.717834617	2.250420002	BP
GO:0010498	proteasomal protein catabolic process	-8.222573178	2.57634135	BP
GO:0010608	post-transcriptional regulation of gene expression	-3.217776213	2.72427587	BP
GO:0010821	regulation of mitochondrion organization	-3.112218223	2.181843588	BP
GO:0015931	nucleobase-containing compound transport	-2.579269333	2.342422681	BP
GO:0016064	immunoglobulin mediated immune response	-2.351591348	2.245512668	BP
GO:0019724	B cell mediated immunity	-2.329190365	2.255272505	BP
GO:0016579	protein deubiquitination	-4.161150909	2.184691431	BP
GO:0017148	negative regulation of translation	-2.485941345	2.222716471	BP
GO:0019693	ribose phosphate metabolic process	-6.866461092	2.618048097	BP
GO:0021782	glial cell development	-2.243741762	2.037426498	BP
GO:0022900	electron transport chain	-7.847711656	2.230448921	BP
GO:0030183	B cell differentiation	-5.075720714	2.11058971	BP
GO:0031396	regulation of protein ubiquitination	-4.254925208	2.324282455	BP
GO:1903322	positive regulation of protein modification by small protein conjugation or removal	-2.748726831	2.1430148	BP
GO:0031398	positive regulation of protein ubiquitination	-3.454632143	2.08278537	BP
GO:0031623	receptor internalization	-2.382358085	1.86332286	BP
GO:0031647	regulation of protein stability	-3.365372618	2.5132176	BP
GO:0033044	regulation of chromosome organization	-2.62384376	2.403120521	BP
GO:0033157	regulation of intracellular protein transport	-4.690369833	2.352182518	BP
GO:0033865	nucleoside bisphosphate metabolic process	-2.910880292	2.089905111	BP
GO:0034032	purine nucleoside bisphosphate metabolic process	-2.910880292	2.089905111	BP
GO:0033875	ribonucleoside bisphosphate metabolic process	-2.910880292	2.089905111	BP
GO:0034248	regulation of amide metabolic process	-8.35753548	2.701567985	BP
GO:0034249	negative regulation of amide metabolic process	-2.669020799	2.276461804	BP
GO:0036503	ERAD pathway	-3.022461592	2	BP
GO:0042176	regulation of protein catabolic process	-4.841637508	2.555094449	BP
GO:0042752	regulation of circadian rhythm	-2.244126609	2.086359831	BP
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	-7.454692884	2.539076099	BP
GO:0043484	regulation of RNA splicing	-1.914018498	2.257678575	BP
GO:0044242	cellular lipid catabolic process	-1.94217739	2.28780173	BP
GO:0044272	sulfur compound biosynthetic process	-1.806406144	2.195899652	BP
GO:0045333	cellular respiration	-10	2.285557309	BP
GO:0045446	endothelial cell differentiation	-3.647858581	1.934498451	BP
GO:0045732	positive regulation of protein catabolic process	-3.395016179	2.324282455	BP
GO:0046390	ribose phosphate biosynthetic process	-5.798602876	2.344392274	BP
GO:0046434	organophosphate catabolic process	-2.314890384	2.195899652	BP
GO:0046683	response to organophosphorus	-3.073052588	2.10720997	BP
GO:0046822	regulation of nucleocytoplasmic transport	-2.718423291	2.037426498	BP
GO:0048193	Golgi vesicle transport	-5.557520231	2.46834733	BP
GO:0048284	organelle fusion	-2.867687089	2.10720997	BP
GO:0050658	RNA transport	-2.748468067	2.204119983	BP
GO:0050657	nucleic acid transport	-2.748468067	2.204119983	BP
GO:0051028	mRNA transport	-1.956436663	2.113943352	BP
GO:0051236	establishment of RNA localization	-2.69416626	2.209515015	BP
GO:0050671	positive regulation of lymphocyte proliferation	-3.58092346	2.164352856	BP
GO:0032946	positive regulation of mononuclear cell proliferation	-3.321485248	2.173186268	BP
GO:0042129	regulation of T cell proliferation	-3.204334917	2.257678575	BP
GO:0050821	protein stabilization	-7.358525889	2.330413773	BP
GO:0050864	regulation of B cell activation	-5.844663963	2.344392274	BP
GO:0050906	detection of stimulus involved in sensory perception	-1.791012593	2.749736316	BP
GO:0051169	nuclear transport	-2.603359112	2.390935107	BP
GO:0051928	positive regulation of calcium ion transport	-2.832861756	2.11058971	BP
GO:0055086	nucleobase-containing small molecule metabolic process	-8.853871964	2.779596491	BP
GO:0070646	protein modification by small protein removal	-3.42157054	2.23299611	BP

GO:0072009	nephron epithelium development	-3.647858581	2.021189299	BP
GO:0072594	establishment of protein localization to organelle	-8.058488567	2.526339277	BP
GO:0090150	establishment of protein localization to membrane	-2.24636278	2.328379603	BP
GO:0090316	positive regulation of intracellular protein transport	-2.511287017	2.193124598	BP
GO:0090501	RNA phosphodiester bond hydrolysis	-2.261981579	2.235528447	BP
GO:0098659	inorganic cation import across plasma membrane	-2.337341778	1.995635195	BP
GO:0099587	inorganic ion import across plasma membrane	-2.337341778	1.995635195	BP
GO:1901136	carbohydrate derivative catabolic process	-2.064596254	2.243038049	BP
GO:1903008	organelle disassembly	-3.852967259	2.017033339	BP
GO:1903038	negative regulation of leukocyte cell-cell adhesion	-3.356414148	2.136720567	BP
GO:0050868	negative regulation of T cell activation	-2.999992318	2.100370545	BP
GO:1903050	regulation of proteolysis involved in protein catabolic process	-3.664582893	2.357934847	BP
GO:1903320	regulation of protein modification by small protein conjugation or removal	-4.345823458	2.403120521	BP
GO:1905037	autophagosome organization	-2.234694186	1.913813852	BP
GO:2000058	regulation of ubiquitin-dependent protein catabolic process	-3.46358424	2.240549248	BP
GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	-3.03867196	2.155336037	BP
GO:0061136	regulation of proteasomal protein catabolic process	-2.866193013	2.285557309	BP
GO:1901800	positive regulation of proteasomal protein catabolic process	-2.111277729	2.075546961	BP
GO:1903052	positive regulation of proteolysis involved in protein catabolic process	-3.0294192	2.1430148	BP
GO:2000060	positive regulation of ubiquitin-dependent protein catabolic process	-2.697925056	2.056904851	BP
GO:2000112	regulation of cellular macromolecule biosynthetic process	-7.37161107	2.720985744	BP
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	-2.660991865	2.269512944	BP
GO:2000278	regulation of DNA biosynthetic process	-2.031427246	2.100370545	BP
GO:2001242	regulation of intrinsic apoptotic signaling pathway	-2.460978887	2.220108088	BP
GO:0000151	ubiquitin ligase complex	-3.180244214	2.498310554	CC
GO:0000228	nuclear chromosome	-2.077992528	2.367355921	CC
GO:0000781	chromosome, telomeric region	-2.598464459	2.187520721	CC
GO:0005681	spliceosomal complex	-8.756961951	2.29666519	CC
GO:0005759	mitochondrial matrix	-10	2.685741739	CC
GO:0005776	autophagosome	-1.720314198	2.075546961	CC
GO:0005777	peroxisome	-6.294992041	2.167317335	CC
GO:0016607	nuclear speck	-4.809668302	2.621176282	CC
GO:0019867	outer membrane	-8.832682665	2.378397901	CC
GO:0030175	filopodium	-1.76291657	2.037426498	CC
GO:0030672	synaptic vesicle membrane	-10	2.08278537	CC
GO:0030863	cortical cytoskeleton	-3.262665348	2.037426498	CC
GO:0030880	RNA polymerase complex	-2.647750451	2.093421685	CC
GO:0031461	cullin-RING ubiquitin ligase complex	-2.34448429	2.269512944	CC
GO:0031965	nuclear membrane	-2.941484825	2.487138375	CC
GO:0031968	organelle outer membrane	-8.787812396	2.374748346	CC
GO:0005741	mitochondrial outer membrane	-7.373659633	2.320146286	CC
GO:0042579	microbody	-6.294992041	2.167317335	CC
GO:0061695	transferase complex, transferring phosphorus-containing groups	-2.585446096	2.496929648	CC
GO:0099501	exocytic vesicle membrane	-10	2.08278537	CC
GO:1902911	protein kinase complex	-2.168819456	2.176091259	CC
GO:1905368	peptidase complex	-7.761953897	2.10720997	CC
GO:1990204	oxidoreductase complex	-9.978810701	2.120573931	CC
GO:0000149	SNARE binding	-2.916294353	2.049218023	MF
GO:0000287	magnesium ion binding	-2.575914598	2.354108439	MF
GO:0001653	peptide receptor activity	-2.235315082	2.187520721	MF
GO:0003697	single-stranded DNA binding	-2.788527305	2.103803721	MF
GO:0003729	mRNA binding	-2.640941087	2.509202522	MF
GO:0004540	ribonuclease activity	-1.919350143	2.136720567	MF
GO:0004866	endopeptidase inhibitor activity	-2.928137068	2.26245109	MF
GO:0005262	calcium channel activity	-5.863279433	2.071882007	MF
GO:0005267	potassium channel activity	-8	2.096910013	MF
GO:0005506	iron ion binding	-2.798209042	2.187520721	MF
GO:0008234	cysteine-type peptidase activity	-2.245531316	2.283301229	MF
GO:0008528	G protein-coupled peptide receptor activity	-2.18229282	2.170261715	MF
GO:0009055	electron transfer activity	-7.756961951	2.075546961	MF
GO:0015078	proton transmembrane transporter activity	-5.249491605	2.152288344	MF
GO:0015399	primary active transmembrane transporter activity	-3.650812694	2.220108088	MF
GO:0016247	channel regulator activity	-3.683637363	2.164352856	MF
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	-3.811280866	2.1430148	MF
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP	-3.665604688	2.11058971	MF

	as acceptor			
GO:0016779	nucleotidyltransferase activity	-1.944399047	2.187520721	MF
GO:0016829	lyase activity	-2.827551578	2.309630167	MF
GO:0016853	isomerase activity	-4.437707136	2.230448921	MF
GO:0016887	ATP hydrolysis activity	-3.253514216	2.551449998	MF
GO:0022853	active ion transmembrane transporter activity	-1.756524903	2.456366033	MF
GO:0031072	heat shock protein binding	-3.441535154	2.096910013	MF
GO:0031625	ubiquitin protein ligase binding	-6.801342913	2.477121255	MF
GO:0032182	ubiquitin-like protein binding	-2.33045443	2.08278537	MF
GO:0045182	translation regulator activity	-4.083019953	2.161368002	MF
GO:0046906	tetrapyrrole binding	-2.619892108	2.178976947	MF
GO:0047485	protein N-terminus binding	-3.853612373	2.029383778	MF
GO:0061630	ubiquitin protein ligase activity	-2.235428584	2.514547753	MF
GO:0061659	ubiquitin-like protein ligase activity	-3.147219134	2.530199698	MF
GO:0099094	ligand-gated cation channel activity	-6.012333735	2.075546961	MF
GO:0099106	ion channel regulator activity	-3.553932178	2.149219113	MF
GO:0140098	catalytic activity, acting on RNA	-3.536647202	2.600972896	MF

9.1.3. Intersecting GO Terms for All Male Tissues, not Found in Cell Samples

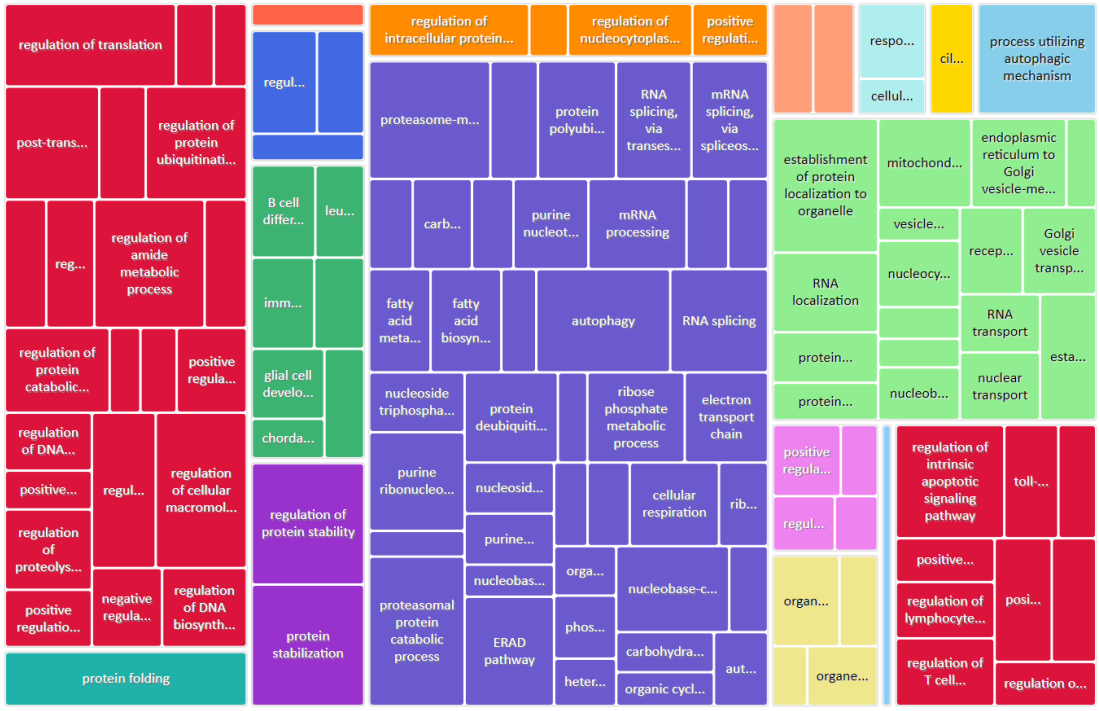

Figure 108 – Enriched BP terms tree plot found in male tissues and not in cells

Figure 109 – Enriched CC terms tree plot found male tissues and not in cells

Figure 110 – Enriched MF terms tree plot found in male tissues and not in cells

Table 31 - Enriched GO terms found in male tissues and not in cells

TermID	Name	Value	LogSize	Term
GO:0000018	regulation of DNA recombination	-2.215138234	2.127104798	BP
GO:0000045	autophagosome assembly	-3.927320911	1.875061263	BP
GO:0000082	G1/S transition of mitotic cell cycle	-3.422632405	1.934498451	BP
GO:0000209	protein polyubiquitination	-6.441291429	2.359835482	BP
GO:0000375	RNA splicing, via transesterification reactions	-6.225483034	2.387389826	BP

GO:0000398	mRNA splicing, via spliceosome	-6.312471039	2.380211242	BP
GO:0000377	RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	-6.312471039	2.380211242	BP
GO:0000723	telomere maintenance	-2.753448603	1.986771734	BP
GO:0001909	leukocyte mediated cytotoxicity	-3.289896225	1.799340549	BP
GO:0002224	toll-like receptor signaling pathway	-4.389339837	1.633468456	BP
GO:0002377	immunoglobulin production	-4.162411562	1.672097858	BP
GO:0002703	regulation of leukocyte mediated immunity	-3.065773537	2.380211242	BP
GO:0002705	positive regulation of leukocyte mediated immunity	-3.100313289	2.1430148	BP
GO:0002706	regulation of lymphocyte mediated immunity	-3.981083297	2.247973266	BP
GO:0005975	carbohydrate metabolic process	-3.748401483	2.675778342	BP
GO:0006006	glucose metabolic process	-2.622363904	2.064457989	BP
GO:0006164	purine nucleotide biosynthetic process	-4.735182177	2.371067862	BP
GO:0009152	purine ribonucleotide biosynthetic process	-2.892695376	2.301029996	BP
GO:0009260	ribonucleotide biosynthetic process	-3.942744524	2.330413773	BP
GO:0006282	regulation of DNA repair	-1.926896574	2.33243846	BP
GO:0006397	mRNA processing	-6.20481541	2.652246341	BP
GO:0006401	RNA catabolic process	-2.641601144	2.214843848	BP
GO:0006402	mRNA catabolic process	-2.511020632	2.10720997	BP
GO:0000956	nuclear-transcribed mRNA catabolic process	-1.994037268	2.017033339	BP
GO:0006403	RNA localization	-5.950781977	2.26245109	BP
GO:0006417	regulation of translation	-10	2.64738297	BP
GO:0006457	protein folding	-10	2.346352974	BP
GO:0006605	protein targeting	-3.768281189	2.394451681	BP
GO:0006612	protein targeting to membrane	-2.794370209	1.963787827	BP
GO:0006631	fatty acid metabolic process	-4.443697499	2.514547753	BP
GO:0006633	fatty acid biosynthetic process	-5.124938737	2.079181246	BP
GO:0006661	phosphatidylinositol biosynthetic process	-2.5296484	2.117271296	BP
GO:0006839	mitochondrial transport	-5.863279433	2.170261715	BP
GO:0006888	endoplasmic reticulum to Golgi vesicle-mediated transport	-5.978810701	2.130333768	BP
GO:0006892	post-Golgi vesicle-mediated transport	-1.874161486	1.991226076	BP
GO:0006906	vesicle fusion	-1.924029887	2.012837225	BP
GO:0090174	organelle membrane fusion	-1.863522909	2.021189299	BP
GO:0006913	nucleocytoplasmic transport	-3.848220393	2.390935107	BP
GO:0051168	nuclear export	-2.11720627	2.123851641	BP
GO:0006914	autophagy	-9.698970004	2.495544338	BP
GO:0006997	nucleus organization	-2.568736969	2.167317335	BP
GO:0007030	Golgi organization	-1.520894525	2.178976947	BP
GO:0007034	vacuolar transport	-1.837008754	2.212187604	BP
GO:0007041	lysosomal transport	-1.65403615	2.113943352	BP
GO:0008380	RNA splicing	-7.079354999	2.563481085	BP
GO:0009141	nucleoside triphosphate metabolic process	-4.008330993	2.320146286	BP
GO:0009150	purine ribonucleotide metabolic process	-6.605548319	2.587710965	BP
GO:0009259	ribonucleotide metabolic process	-2.629324732	2.608526034	BP
GO:0009451	RNA modification	-1.72299291	2.250420002	BP
GO:0009566	fertilization	-2.181656616	2.283301229	BP
GO:0010498	proteasomal protein catabolic process	-10	2.57634135	BP
GO:0010608	post-transcriptional regulation of gene expression	-7.537602002	2.72427587	BP
GO:0010821	regulation of mitochondrion organization	-5	2.181843588	BP
GO:0015931	nucleobase-containing compound transport	-3.054578439	2.342422681	BP
GO:0016064	immunoglobulin mediated immune response	-3.329785839	2.245512668	BP
GO:0019724	B cell mediated immunity	-3.251100903	2.255272505	BP
GO:0016579	protein deubiquitination	-5.89279003	2.184691431	BP
GO:0017148	negative regulation of translation	-3.712295624	2.222716471	BP
GO:0019439	aromatic compound catabolic process	-1.86098623	2.544068044	BP
GO:0019693	ribose phosphate metabolic process	-6.154281982	2.618048097	BP
GO:0021782	glial cell development	-3.713830403	2.037426498	BP
GO:0022900	electron transport chain	-5.26520017	2.230448921	BP
GO:0030183	B cell differentiation	-4.278189385	2.11058971	BP
GO:0031396	regulation of protein ubiquitination	-8.07211659	2.324282455	BP
GO:1903322	positive regulation of protein modification by small protein conjugation or removal	-3.547243745	2.1430148	BP
GO:0031398	positive regulation of protein ubiquitination	-3.752864086	2.08278537	BP
GO:0031623	receptor internalization	-3.873966912	1.86332286	BP
GO:0031647	regulation of protein stability	-10	2.5132176	BP

GO:0032411	positive regulation of transporter activity	-1.545083904	2.100370545	BP
GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	-4.35753548	2.155336037	BP
GO:0061136	regulation of proteasomal protein catabolic process	-4.484126156	2.285557309	BP
GO:1901800	positive regulation of proteasomal protein catabolic process	-3.139580022	2.075546961	BP
GO:2000058	regulation of ubiquitin-dependent protein catabolic process	-4.718966633	2.240549248	BP
GO:2000060	positive regulation of ubiquitin-dependent protein catabolic process	-2.991981683	2.056904851	BP
GO:0033044	regulation of chromosome organization	-3.625942653	2.403120521	BP
GO:0033157	regulation of intracellular protein transport	-6.401209493	2.352182518	BP
GO:0033865	nucleoside bisphosphate metabolic process	-3.202977871	2.089905111	BP
GO:0034032	purine nucleoside bisphosphate metabolic process	-3.202977871	2.089905111	BP
GO:0033875	ribonucleoside bisphosphate metabolic process	-3.202977871	2.089905111	BP
GO:0034248	regulation of amide metabolic process	-10	2.701567985	BP
GO:0034249	negative regulation of amide metabolic process	-3.827755973	2.276461804	BP
GO:0034655	nucleobase-containing compound catabolic process	-2.010099782	2.457881897	BP
GO:0036503	ERAD pathway	-6.847711656	2	BP
GO:0042129	regulation of T cell proliferation	-4.787812396	2.257678575	BP
GO:0032946	positive regulation of mononuclear cell proliferation	-4.389339837	2.173186268	BP
GO:0050671	positive regulation of lymphocyte proliferation	-4.653647026	2.164352856	BP
GO:0042176	regulation of protein catabolic process	-6.28567024	2.555094449	BP
GO:0043009	chordate embryonic development	-2.095255301	2.812913357	BP
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic process	-10	2.539076099	BP
GO:0043467	regulation of generation of precursor metabolites and energy	-1.84559329	2.120573931	BP
GO:0043484	regulation of RNA splicing	-2.171586544	2.257678575	BP
GO:0044270	cellular nitrogen compound catabolic process	-1.962838718	2.526339277	BP
GO:0044272	sulfur compound biosynthetic process	-2.477176405	2.195899652	BP
GO:0044770	cell cycle phase transition	-3.375082355	2.26245109	BP
GO:0045089	positive regulation of innate immune response	-5.061980903	2.392696953	BP
GO:0045333	cellular respiration	-5.235823868	2.285557309	BP
GO:0045446	endothelial cell differentiation	-3.178945313	1.934498451	BP
GO:0045732	positive regulation of protein catabolic process	-4.208309351	2.324282455	BP
GO:0045931	positive regulation of mitotic cell cycle	-1.963849609	2.079181246	BP
GO:0046390	ribose phosphate biosynthetic process	-2.874791246	2.344392274	BP
GO:0046434	organophosphate catabolic process	-2.154741408	2.195899652	BP
GO:0046488	phosphatidylinositol metabolic process	-2.747416181	2.198657087	BP
GO:0046683	response to organophosphorus	-3.840350486	2.10720997	BP
GO:0046700	heterocycle catabolic process	-2.06144342	2.523746467	BP
GO:0046822	regulation of nucleocytoplasmic transport	-4.974694135	2.037426498	BP
GO:0048193	Golgi vesicle transport	-4.521433504	2.46834733	BP
GO:0048284	organelle fusion	-3.133572816	2.10720997	BP
GO:0050658	RNA transport	-3.305854122	2.204119983	BP
GO:0050657	nucleic acid transport	-3.305854122	2.204119983	BP
GO:0051028	mRNA transport	-2.355609305	2.113943352	BP
GO:0051236	establishment of RNA localization	-2.970418291	2.209515015	BP
GO:0050821	protein stabilization	-10	2.330413773	BP
GO:0050864	regulation of B cell activation	-3.933973402	2.344392274	BP
GO:0051052	regulation of DNA metabolic process	-3.56235756	2.729974286	BP
GO:0051054	positive regulation of DNA metabolic process	-2.390240716	2.482873584	BP
GO:0051169	nuclear transport	-3.848220393	2.390935107	BP
GO:0055086	nucleobase-containing small molecule metabolic process	-6.812479279	2.779596491	BP
GO:0060294	cilium movement involved in cell motility	-3.781572638	2.093421685	BP
GO:0061919	process utilizing autophagic mechanism	-9.698970004	2.495544338	BP
GO:0070646	protein modification by small protein removal	-2.337725671	2.23299611	BP
GO:0071466	cellular response to xenobiotic stimulus	-1.797409903	2.26245109	BP
GO:0072594	establishment of protein localization to organelle	-10	2.526339277	BP
GO:0090150	establishment of protein localization to membrane	-5.005682847	2.328379603	BP
GO:0090316	positive regulation of intracellular protein transport	-3.023319431	2.193124598	BP
GO:1901136	carbohydrate derivative catabolic process	-2.746120751	2.243038049	BP
GO:1901361	organic cyclic compound catabolic process	-2.313497866	2.583198774	BP
GO:1901987	regulation of cell cycle phase transition	-2.532717843	2.631443769	BP
GO:1901988	negative regulation of cell cycle phase transition	-1.755099705	2.346352974	BP
GO:1901991	negative regulation of mitotic cell cycle phase transition	-1.72521495	2.212187604	BP
GO:1901989	positive regulation of cell cycle phase transition	-3.609877106	2.049218023	BP
GO:1902115	regulation of organelle assembly	-2.240285398	2.369215857	BP
GO:1903008	organelle disassembly	-4.474955193	2.017033339	BP
GO:1903038	negative regulation of leukocyte cell-cell adhesion	-2.592870077	2.136720567	BP

GO:1903050	regulation of proteolysis involved in protein catabolic process	-4.943095149	2.357934847	BP
GO:1903052	positive regulation of proteolysis involved in protein catabolic process	-3.52215532	2.1430148	BP
GO:1903320	regulation of protein modification by small protein conjugation or removal	-7.003926346	2.403120521	BP
GO:1905037	autophagosome organization	-2.799286392	1.913813852	BP
GO:2000112	regulation of cellular macromolecule biosynthetic process	-10	2.720985744	BP
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	-3.916174128	2.269512944	BP
GO:2000278	regulation of DNA biosynthetic process	-4.732828272	2.100370545	BP
GO:2001020	regulation of response to DNA damage stimulus	-3.209053704	2.496929648	BP
GO:2001242	regulation of intrinsic apoptotic signaling pathway	-8.798602876	2.220108088	BP
GO:0000151	ubiquitin ligase complex	-7.950781977	2.498310554	CC
GO:0000781	chromosome, telomeric region	-3.907744675	2.187520721	CC
GO:0005681	spliceosomal complex	-7.380906669	2.29666519	CC
GO:0005741	mitochondrial outer membrane	-6.701146924	2.320146286	CC
GO:0031968	organelle outer membrane	-6.061480275	2.374748346	CC
GO:0005759	mitochondrial matrix	-9.764471553	2.685741739	CC
GO:0005776	autophagosome	-1.834820199	2.075546961	CC
GO:0005777	peroxisome	-3.273776058	2.167317335	CC
GO:0009898	cytoplasmic side of plasma membrane	-1.769588275	2.240549248	CC
GO:0016607	nuclear speck	-2.826045019	2.621176282	CC
GO:0019867	outer membrane	-6.258060922	2.378397901	CC
GO:0030672	synaptic vesicle membrane	-3.763627526	2.08278537	CC
GO:0030863	cortical cytoskeleton	-1.903965883	2.037426498	CC
GO:0030880	RNA polymerase complex	-2.507920277	2.093421685	CC
GO:0031461	cullin-RING ubiquitin ligase complex	-3.797746517	2.269512944	CC
GO:0031965	nuclear membrane	-4.297569464	2.487138375	CC
GO:0042579	microbody	-3.273776058	2.167317335	CC
GO:0044853	plasma membrane raft	-2.226046126	2.06069784	CC
GO:0061695	transferase complex, transferring phosphorus-containing groups	-3.30436325	2.496929648	CC
GO:0098562	cytoplasmic side of membrane	-1.600864952	2.315970345	CC
GO:0099501	exocytic vesicle membrane	-3.763627526	2.08278537	CC
GO:1902911	protein kinase complex	-2.182276486	2.176091259	CC
GO:1905368	peptidase complex	-6.739928612	2.10720997	CC
GO:1990204	oxidoreductase complex	-2.9421374	2.120573931	CC
GO:0000149	SNARE binding	-2.791219627	2.049218023	MF
GO:0000287	magnesium ion binding	-3.826657493	2.354108439	MF
GO:0003697	single-stranded DNA binding	-2.511847834	2.103803721	MF
GO:0003729	mRNA binding	-3.574350252	2.509202522	MF
GO:0004842	ubiquitin-protein transferase activity	-5.110138279	2.64738297	MF
GO:0061630	ubiquitin protein ligase activity	-2.504081637	2.514547753	MF
GO:0061659	ubiquitin-like protein ligase activity	-2.355182059	2.530199698	MF
GO:0005198	structural molecule activity	-8.730487056	2.893206753	MF
GO:0005267	potassium channel activity	-5.586700236	2.096910013	MF
GO:0005506	iron ion binding	-1.959007898	2.187520721	MF
GO:0008234	cysteine-type peptidase activity	-2.501213755	2.283301229	MF
GO:0008528	G protein-coupled peptide receptor activity	-2.753558841	2.170261715	MF
GO:0009055	electron transfer activity	-4.095284455	2.075546961	MF
GO:0015103	inorganic anion transmembrane transporter activity	-1.921618422	2.206825876	MF
GO:0015399	primary active transmembrane transporter activity	-2.135972624	2.220108088	MF
GO:0016247	channel regulator activity	-2.076782438	2.164352856	MF
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	-3.778434071	2.1430148	MF
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	-3.116501939	2.11058971	MF
GO:0016746	acyltransferase activity	-2.037223224	2.409933123	MF
GO:0016747	acyltransferase activity, transferring groups other than amino-acyl groups	-2.252603654	2.361727836	MF
GO:0016829	lyase activity	-2.136822247	2.309630167	MF
GO:0016853	isomerase activity	-4.91721463	2.230448921	MF
GO:0016887	ATP hydrolysis activity	-3.684195689	2.551449998	MF
GO:0019208	phosphatase regulator activity	-2.516007318	2.06069784	MF
GO:0019787	ubiquitin-like protein transferase activity	-4.886056648	2.671172843	MF
GO:0019902	phosphatase binding	-2.186124675	2.292256071	MF
GO:0031072	heat shock protein binding	-10	2.096910013	MF
GO:0031625	ubiquitin protein ligase binding	-10	2.477121255	MF
GO:0032182	ubiquitin-like protein binding	-2.647521526	2.08278537	MF
GO:0042393	histone binding	-1.651789553	2.411619706	MF
GO:0045182	translation regulator activity	-7.391473966	2.161368002	MF

GO:0046906	tetrapyrrole binding	-1.904072645	2.178976947	MF
GO:0047485	protein N-terminus binding	-3.601096767	2.029383778	MF
GO:0099106	ion channel regulator activity	-2.052265973	2.149219113	MF
GO:0140098	catalytic activity, acting on RNA	-2.190153739	2.600972896	MF
GO:1901981	phosphatidylinositol phosphate binding	-1.932550998	2.260071388	MF

9.1.4. Intersecting GO Terms Specific to Female Tissues

Figure 111 – Enriched BP terms tree plot only found in female tissues.

Figure 112 – Enriched CC terms tree plot only found in female tissues.

Figure 113 – Enriched MF terms tree plot only found in female tissues.

Table 32 - Enriched GO terms only found in female tissues.

TermID	Name	Value	LogSize	Term
GO:0006310	DNA recombination	-1.917863759	2.385606274	BP
GO:0009308	amine metabolic process	-1.719757793	1.982271233	BP
GO:0009408	response to heat	-2.889872181	2.021189299	BP
GO:0042752	regulation of circadian rhythm	-2.244126609	2.086359831	BP
GO:0044242	cellular lipid catabolic process	-1.94217739	2.28780173	BP
GO:0050868	negative regulation of T cell activation	-2.999992318	2.100370545	BP
GO:0050906	detection of stimulus involved in sensory perception	-1.791012593	2.749736316	BP
GO:0051928	positive regulation of calcium ion transport	-2.832861756	2.11058971	BP
GO:0072009	nephron epithelium development	-3.647858581	2.021189299	BP
GO:0090501	RNA phosphodiester bond hydrolysis	-2.261981579	2.235528447	BP
GO:0098659	inorganic cation import across plasma membrane	-2.337341778	1.995635195	BP
GO:0099587	inorganic ion import across plasma membrane	-2.337341778	1.995635195	BP
GO:0000228	nuclear chromosome	-2.077992528	2.367355921	CC
GO:0030175	filopodium	-1.76291657	2.037426498	CC
GO:0001653	peptide receptor activity	-2.235315082	2.187520721	MF
GO:0004540	ribonuclease activity	-1.919350143	2.136720567	MF
GO:0004866	endopeptidase inhibitor activity	-2.928137068	2.26245109	MF
GO:0005262	calcium channel activity	-5.863279433	2.071882007	MF
GO:0015078	proton transmembrane transporter activity	-5.249491605	2.152288344	MF
GO:0016779	nucleotidyltransferase activity	-1.944399047	2.187520721	MF
GO:0022853	active ion transmembrane transporter activity	-1.756524903	2.456366033	MF
GO:0099094	ligand-gated cation channel activity	-6.012333735	2.075546961	MF

9.1.5. Intersecting GO Terms Specific to Male Tissues

Figure 114 – Enriched BP terms tree plot only found in male tissues.

Figure 115 – Enriched CC terms tree plot only found in male tissues.

Figure 116 – Enriched MF terms tree plot only found in male tissues.

Table 33 - Enriched GO terms only found in male tissues.

TermID	Name	Value	LogSize	Term
GO:0000082	G1/S transition of mitotic cell cycle	-3.422632405	1.934498451	BP
GO:0000956	nuclear-transcribed mRNA catabolic process	-1.994037268	2.017033339	BP
GO:0002224	toll-like receptor signaling pathway	-4.389339837	1.633468456	BP
GO:0005975	carbohydrate metabolic process	-3.748401483	2.675778342	BP
GO:0006282	regulation of DNA repair	-1.926896574	2.33243846	BP
GO:0006892	post-Golgi vesicle-mediated transport	-1.874161486	1.991226076	BP
GO:0006914	autophagy	-9.698970004	2.495544338	BP
GO:0006997	nucleus organization	-2.568736969	2.167317335	BP
GO:0007030	Golgi organization	-1.520894525	2.178976947	BP
GO:0007041	lysosomal transport	-1.65403615	2.113943352	BP
GO:0009566	fertilization	-2.181656616	2.283301229	BP
GO:0019439	aromatic compound catabolic process	-1.86098623	2.544068044	BP
GO:0032411	positive regulation of transporter activity	-1.545083904	2.100370545	BP
GO:0034655	nucleobase-containing compound catabolic process	-2.010099782	2.457881897	BP
GO:0043009	chordate embryonic development	-2.095255301	2.812913357	BP
GO:0043467	regulation of generation of precursor metabolites and energy	-1.84559329	2.120573931	BP
GO:0044270	cellular nitrogen compound catabolic process	-1.962838718	2.526339277	BP
GO:0044770	cell cycle phase transition	-3.375082355	2.26245109	BP
GO:0045089	positive regulation of innate immune response	-5.061980903	2.392696953	BP
GO:0045931	positive regulation of mitotic cell cycle	-1.963849609	2.079181246	BP
GO:0046488	phosphatidylinositol metabolic process	-2.747416181	2.198657087	BP
GO:0046700	heterocycle catabolic process	-2.06144342	2.523746467	BP
GO:0051052	regulation of DNA metabolic process	-3.56235756	2.729974286	BP
GO:0051054	positive regulation of DNA metabolic process	-2.390240716	2.482873584	BP
GO:0060294	cilium movement involved in cell motility	-3.781572638	2.093421685	BP
GO:0061919	process utilizing autophagic mechanism	-9.698970004	2.495544338	BP
GO:0071466	cellular response to xenobiotic stimulus	-1.797409903	2.26245109	BP
GO:1901361	organic cyclic compound catabolic process	-2.313497866	2.583198774	BP
GO:1901987	regulation of cell cycle phase transition	-2.532717843	2.631443769	BP
GO:1901988	negative regulation of cell cycle phase transition	-1.755099705	2.346352974	BP
GO:1901991	negative regulation of mitotic cell cycle phase transition	-1.72521495	2.212187604	BP
GO:1901989	positive regulation of cell cycle phase transition	-3.609877106	2.049218023	BP
GO:1902115	regulation of organelle assembly	-2.240285398	2.369215857	BP

GO:2001020	regulation of response to DNA damage stimulus	-3.209053704	2.496929648	BP
GO:0009898	cytoplasmic side of plasma membrane	-1.769588275	2.240549248	CC
GO:0044853	plasma membrane raft	-2.226046126	2.06069784	CC
GO:0098562	cytoplasmic side of membrane	-1.600864952	2.315970345	CC
GO:0004842	ubiquitin-protein transferase activity	-5.110138279	2.64738297	MF
GO:0005198	structural molecule activity	-8.730487056	2.893206753	MF
GO:0015103	inorganic anion transmembrane transporter activity	-1.921618422	2.206825876	MF
GO:0016746	acyltransferase activity	-2.037223224	2.409933123	MF
GO:0016747	acyltransferase activity, transferring groups other than amino-acyl groups	-2.252603654	2.361727836	MF
GO:0019208	phosphatase regulator activity	-2.516007318	2.06069784	MF
GO:0019787	ubiquitin-like protein transferase activity	-4.886056648	2.671172843	MF
GO:0019902	phosphatase binding	-2.186124675	2.292256071	MF
GO:0042393	histone binding	-1.651789553	2.411619706	MF
GO:1901981	phosphatidylinositol phosphate binding	-1.932550998	2.260071388	MF

9.1.6. Intersecting GO Terms for All Astrocyte Samples, not Found in Tissue Samples

Figure 117 – Enriched BP terms tree plot found in astrocytes and not tissues.

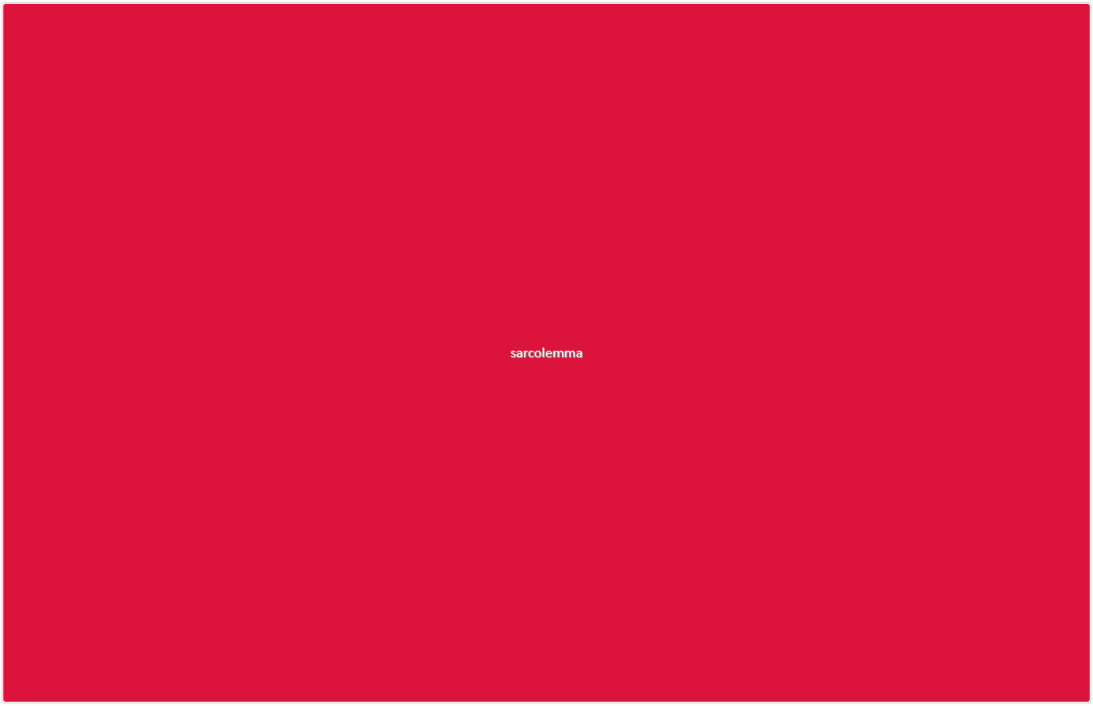

Figure 118 – Enriched CC terms tree plot found in astrocytes and not tissues.

Table 34 - Enriched GO terms found in astrocytes and not tissues.

TermID	Name	Value	LogSize	Term
GO:0002062	chondrocyte differentiation	-6.4225082	1.919078092	BP
GO:0007286	spermatid development	-4.097997109	2.294466226	BP
GO:0007368	determination of left/right symmetry	-2.515364482	2.113943352	BP
GO:0009855	determination of bilateral symmetry	-2.390088396	2.133538908	BP
GO:0007601	visual perception	-2.763685148	2.342422681	BP
GO:0009799	specification of symmetry	-2.113474002	2.136720567	BP
GO:0022412	cellular process involved in reproduction in multicellular organism	-2.254873454	2.632457292	BP
GO:0030203	glycosaminoglycan metabolic process	-3.265399385	2.056904851	BP
GO:0045664	regulation of neuron differentiation	-3.312639923	2.290034611	BP
GO:0048515	spermatid differentiation	-4.268411235	2.309630167	BP
GO:0048706	embryonic skeletal system development	-3.238613653	2.11058971	BP
GO:0071417	cellular response to organonitrogen compound	-5.009661145	2.763427994	BP
GO:0071695	anatomical structure maturation	-4.494850022	2.334453751	BP
GO:0048469	cell maturation	-3.245421769	2.227886705	BP
GO:0098739	import across plasma membrane	-2.350406265	2.198657087	BP
GO:0042383	sarcolemma	-2.126415639	2.136720567	CC

9.1.7. Intersecting GO Terms for All Microglia Samples, not Found in Tissue Samples

Figure 119 – Enriched BP terms tree plot found in microglia and not tissues.

Figure 120 – Enriched CC terms tree plot found in microglia and not tissues.

Figure 121 – Enriched MF terms tree plot found in microglia and not tissues.

Table 35 - Enriched GO terms found in microglia and not tissues.

TermID	Name	Value	LogSize	Term
GO:0000904	cell morphogenesis involved in differentiation	-4.089909454	2.739572344	BP
GO:0001568	blood vessel development	-6.321481621	2.704150517	BP
GO:0001944	vasculature development	-6.369572125	2.721810615	BP
GO:0007268	chemical synaptic transmission	-4.084072788	2.626340367	BP
GO:0098916	anterograde trans-synaptic signaling	-4.084072788	2.626340367	BP
GO:0099537	trans-synaptic signaling	-4.290730039	2.643452676	BP
GO:0007368	determination of left/right symmetry	-5.571865206	2.113943352	BP
GO:0009855	determination of bilateral symmetry	-5.070581074	2.133538908	BP
GO:0007420	brain development	-2.808430559	2.890979597	BP
GO:0007601	visual perception	-2.974314186	2.342422681	BP
GO:0008285	negative regulation of cell population proliferation	-5.450996738	2.846337112	BP
GO:0009799	specification of symmetry	-4.809668302	2.136720567	BP
GO:0030001	metal ion transport	-3.370058689	2.812244697	BP
GO:0030048	actin filament-based movement	-2.134981824	1.995635195	BP
GO:0030203	glycosaminoglycan metabolic process	-4.88941029	2.056904851	BP
GO:0031056	regulation of histone modification	-3.635117155	2.214843848	BP
GO:0032989	cellular component morphogenesis	-4.48148606	2.780317312	BP
GO:0038127	ERBB signaling pathway	-2.486535848	1.799340549	BP
GO:0040008	regulation of growth	-6.131355562	2.788875116	BP
GO:0043244	regulation of protein-containing complex disassembly	-1.805573512	2.10720997	BP
GO:0045664	regulation of neuron differentiation	-2.796394589	2.290034611	BP
GO:0048589	developmental growth	-3.546475148	2.606381365	BP
GO:0048812	neuron projection morphogenesis	-4.978810701	2.677606953	BP
GO:0032990	cell part morphogenesis	-4.614393726	2.703291378	BP
GO:0048858	cell projection morphogenesis	-5.098541679	2.686636269	BP
GO:0120039	plasma membrane bounded cell projection morphogenesis	-5.228412519	2.682145076	BP
GO:0051282	regulation of sequestering of calcium ion	-1.894390229	2.096910013	BP
GO:0051283	negative regulation of sequestering of calcium ion	-1.857770726	1.72427587	BP
GO:0051302	regulation of cell division	-1.744924322	2.278753601	BP
GO:0071363	cellular response to growth factor stimulus	-4.512861625	2.679427897	BP
GO:0071417	cellular response to organonitrogen compound	-2.991502273	2.763427994	BP
GO:0071695	anatomical structure maturation	-2.320617794	2.334453751	BP
GO:0098655	cation transmembrane transport	-2.315782637	2.877371346	BP

GO:0099536	synaptic signaling	-5.931814138	2.668385917	BP
GO:0120035	regulation of plasma membrane bounded cell projection organization	-2.083670427	2.81756537	BP
GO:1901699	cellular response to nitrogen compound	-4.480172006	2.805500858	BP
GO:0005795	Golgi stack	-2.816528784	2.187520721	CC
GO:0030424	axon	-3.35874611	2.808885867	CC
GO:0030425	dendrite	-4.514278574	2.797267541	CC
GO:0042383	sarcolemma	-3.761951535	2.136720567	CC
GO:0097447	dendritic tree	-4.298432015	2.798650645	CC
GO:0098794	postsynapse	-4.468521083	2.821513528	CC
GO:0004672	protein kinase activity	-1.964708225	2.767155866	MF
GO:0008194	UDP-glycosyltransferase activity	-3.130642329	2.164352856	MF
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	-1.936697533	2.190331698	MF
GO:0017171	serine hydrolase activity	-2.549919786	2.301029996	MF
GO:0031406	carboxylic acid binding	-2.426996525	2.26245109	MF

9.1.8. Intersecting GO Terms for All Neuron Samples, not Found in Tissue Samples

Figure 122 – Enriched BP terms tree plot found in neuron and not tissues.

Figure 123 – Enriched CC terms tree plot found in neuron and not tissues.

Figure 124 – Enriched MF terms tree plot found in neuron and not tissues.

Table 36 - Enriched GO terms found in neuron and not tissues.

TermID	Name	Value	LogSize	Term
GO:0007601	visual perception	-2.29921773	2.342422681	BP
GO:0009799	specification of symmetry	-2.600782897	2.136720567	BP
GO:0009855	determination of bilateral symmetry	-2.5268761	2.133538908	BP
GO:0007368	determination of left/right symmetry	-2.065921018	2.113943352	BP
GO:0018107	peptidyl-threonine phosphorylation	-1.518552203	1.851258349	BP

GO:0022412	cellular process involved in reproduction in multicellular organism	-1.593288544	2.632457292	BP
GO:0030038	contractile actin filament bundle assembly	-1.702773938	1.322219295	BP
GO:0030048	actin filament-based movement	-2.859484549	1.995635195	BP
GO:0030203	glycosaminoglycan metabolic process	-3.74790332	2.056904851	BP
GO:0031109	microtubule polymerization or depolymerization	-3.759900535	1.77815125	BP
GO:0031570	DNA integrity checkpoint signaling	-1.896607754	2.056904851	BP
GO:0000077	DNA damage checkpoint signaling	-1.575886177	2.025305865	BP
GO:0038127	ERBB signaling pathway	-2.738148022	1.799340549	BP
GO:0043149	stress fiber assembly	-1.702773938	1.322219295	BP
GO:0043244	regulation of protein-containing complex disassembly	-1.973800798	2.10720997	BP
GO:0045664	regulation of neuron differentiation	-2.174579282	2.290034611	BP
GO:0051225	spindle assembly	-2.777644946	2.068185862	BP
GO:0051302	regulation of cell division	-1.796902257	2.278753601	BP
GO:0061982	meiosis I cell cycle process	-2.396067818	2.11058971	BP
GO:0071695	anatomical structure maturation	-2.865652641	2.334453751	BP
GO:0048469	cell maturation	-2.469636982	2.227886705	BP
GO:0072329	monocarboxylic acid catabolic process	-2.753920963	2.045322979	BP
GO:0098739	import across plasma membrane	-2.054459622	2.198657087	BP
GO:0042383	sarcolemma	-2.279758223	2.136720567	CC
GO:0072686	mitotic spindle	-2.213511396	2.264817823	CC
GO:0005319	lipid transporter activity	-2.423329131	2.235528447	MF
GO:0008514	organic anion transmembrane transporter activity	-2.399070765	2.29666519	MF
GO:0016798	hydrolase activity, acting on glycosyl bonds	-2.02056062	2.1430148	MF
GO:0017171	serine hydrolase activity	-2.33942432	2.301029996	MF
GO:0019842	vitamin binding	-2.422163818	2.173186268	MF
GO:0031406	carboxylic acid binding	-2.663225036	2.26245109	MF
GO:0052689	carboxylic ester hydrolase activity	-2.208699559	2.190331698	MF

9.1.9. Intersecting GO Terms Specific to Astrocyte Samples

Figure 125 – Enriched BP terms tree plot found only in astrocytes.

Table 37 - Enriched GO terms found only in astrocytes.

TermID	Name	Value	LogSize	Term
GO:0002062	chondrocyte differentiation	-6.4225082	1.919078092	BP
GO:0007286	spermatid development	-4.097997109	2.294466226	BP

GO:0048515	spermatid differentiation	-4.268411235	2.309630167	BP
GO:0048706	embryonic skeletal system development	-3.238613653	2.11058971	BP

9.1.10. Intersecting GO Terms Specific to Microglia Samples

Figure 126 – Enriched BP terms tree plot found only in microglia.

Figure 127 – Enriched CC terms tree plot found only in microglia.

Figure 128 – Enriched MF terms tree plot found only in microglia.

Table 38 - Enriched GO terms found only in microglia.

TermID	Name	Value	LogSize	Term
GO:0000904	cell morphogenesis involved in differentiation	-4.089909454	2.739572344	BP
GO:0001568	blood vessel development	-6.321481621	2.704150517	BP
GO:0001944	vasculature development	-6.369572125	2.721810615	BP
GO:0007268	chemical synaptic transmission	-4.084072788	2.626340367	BP
GO:0098916	anterograde trans-synaptic signaling	-4.084072788	2.626340367	BP
GO:0099537	trans-synaptic signaling	-4.290730039	2.643452676	BP
GO:0007420	brain development	-2.808430559	2.890979597	BP
GO:0008285	negative regulation of cell population proliferation	-5.450996738	2.846337112	BP
GO:0030001	metal ion transport	-3.370058689	2.812244697	BP
GO:0031056	regulation of histone modification	-3.635117155	2.214843848	BP
GO:0032989	cellular component morphogenesis	-4.48148606	2.780317312	BP
GO:0040008	regulation of growth	-6.131355562	2.788875116	BP
GO:0048589	developmental growth	-3.546475148	2.606381365	BP
GO:0048812	neuron projection morphogenesis	-4.978810701	2.677606953	BP
GO:0032990	cell part morphogenesis	-4.614393726	2.703291378	BP
GO:0048858	cell projection morphogenesis	-5.098541679	2.686636269	BP
GO:0120039	plasma membrane bounded cell projection morphogenesis	-5.228412519	2.682145076	BP
GO:0051282	regulation of sequestering of calcium ion	-1.894390229	2.096910013	BP
GO:0051283	negative regulation of sequestering of calcium ion	-1.857770726	1.72427587	BP
GO:0071363	cellular response to growth factor stimulus	-4.512861625	2.679427897	BP
GO:0098655	cation transmembrane transport	-2.315782637	2.877371346	BP
GO:0099536	synaptic signaling	-5.931814138	2.668385917	BP
GO:0120035	regulation of plasma membrane bounded cell projection organization	-2.083670427	2.81756537	BP
GO:1901699	cellular response to nitrogen compound	-4.480172006	2.805500858	BP
GO:0005795	Golgi stack	-2.816528784	2.187520721	CC
GO:0030424	axon	-3.35874611	2.808885867	CC
GO:0030425	dendrite	-4.514278574	2.797267541	CC
GO:0097447	dendritic tree	-4.298432015	2.798650645	CC
GO:0098794	postsynapse	-4.468521083	2.821513528	CC
GO:0004672	protein kinase activity	-1.964708225	2.767155866	MF
GO:0008194	UDP-glycosyltransferase activity	-3.130642329	2.164352856	MF
GO:0008757	S-adenosylmethionine-dependent methyltransferase activity	-1.936697533	2.190331698	MF

9.1.11. Intersecting GO Terms Specific to Neuron Samples

Figure 129 – Enriched BP terms tree plot found only in neurons.

Figure 130 – Enriched CC terms tree plot found only in neurons

Figure 131 – Enriched MF terms tree plot found only in neurons

Table 39 - Enriched GO terms found only in neurons

TermID	Name	Value	LogSize	Term
GO:0018107	peptidyl-threonine phosphorylation	-1.518552203	1.851258349	BP
GO:0030038	contractile actin filament bundle assembly	-1.702773938	1.322219295	BP
GO:0031109	microtubule polymerization or depolymerization	-3.759900535	1.77815125	BP
GO:0031570	DNA integrity checkpoint signaling	-1.896607754	2.056904851	BP
GO:0000077	DNA damage checkpoint signaling	-1.575886177	2.025305865	BP
GO:0043149	stress fiber assembly	-1.702773938	1.322219295	BP
GO:0051225	spindle assembly	-2.777644946	2.068185862	BP
GO:0061982	meiosis I cell cycle process	-2.396067818	2.11058971	BP
GO:0072329	monocarboxylic acid catabolic process	-2.753920963	2.045322979	BP
GO:0072686	mitotic spindle	-2.213511396	2.264817823	CC
GO:0005319	lipid transporter activity	-2.423329131	2.235528447	MF
GO:0008514	organic anion transmembrane transporter activity	-2.399070765	2.29666519	MF
GO:0016798	hydrolase activity, acting on glycosyl bonds	-2.02056062	2.1430148	MF
GO:0019842	vitamin binding	-2.422163818	2.173186268	MF
GO:0052689	carboxylic ester hydrolase activity	-2.208699559	2.190331698	MF

9.1.12. Intersecting GO Terms found in Neuron and Microglia Samples, but not in Astrocyte Samples

Figure 132 – Enriched BP terms tree plot common only in neurons and microglia

Figure 133 – Enriched MF terms tree plot common only in neurons and microglia

Table 40 - Enriched GO terms tree plot common only in neurons and microglia

TermID	Name	Value	LogSize	Term
GO:0030048	actin filament-based movement	-2.36098314	1.995635195	BP

GO:0038127	ERBB signaling pathway	-2.594369512	1.799340549	BP
GO:0043244	regulation of protein-containing complex disassembly	-1.88159206	2.10720997	BP
GO:0051302	regulation of cell division	-1.770136139	2.278753601	BP
GO:0017171	serine hydrolase activity	-2.432041996	2.301029996	MF
GO:0031406	carboxylic acid binding	-2.529243309	2.26245109	MF

9.2. Complete Distance Matrices

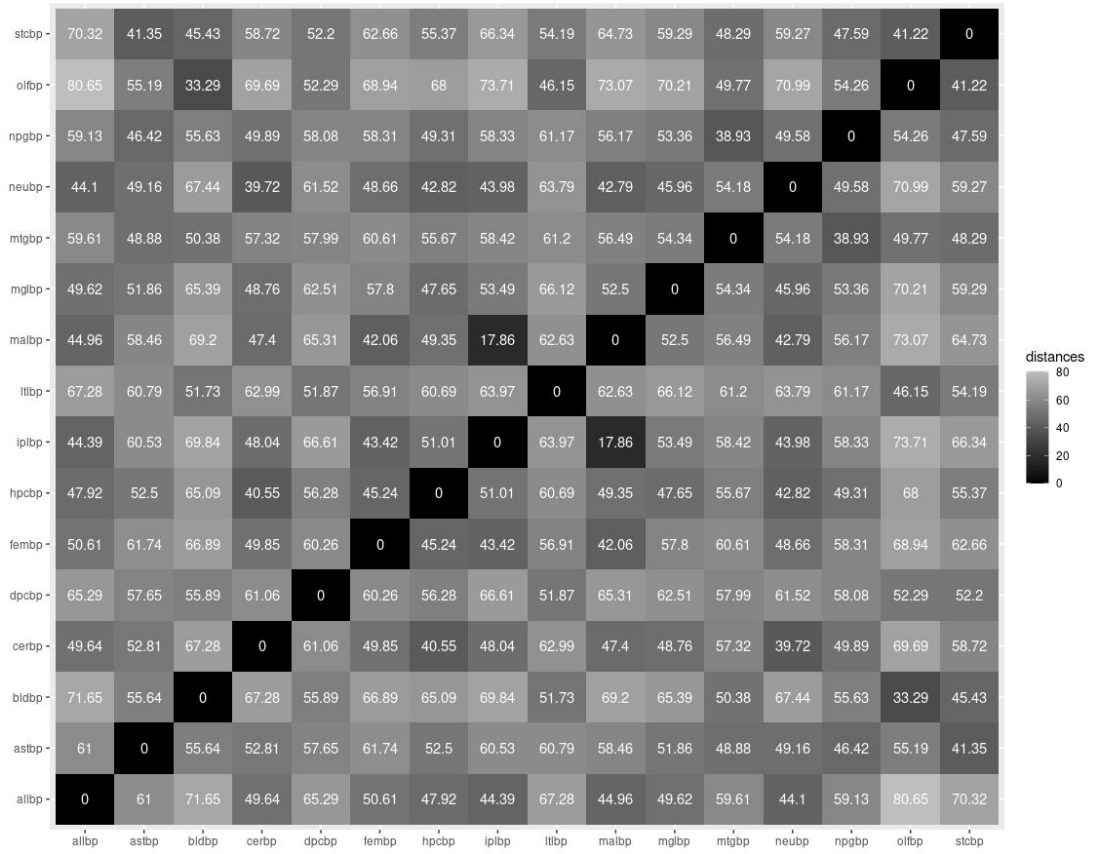

Figure 134 – Distance matrix of all sample groups, built upon BP

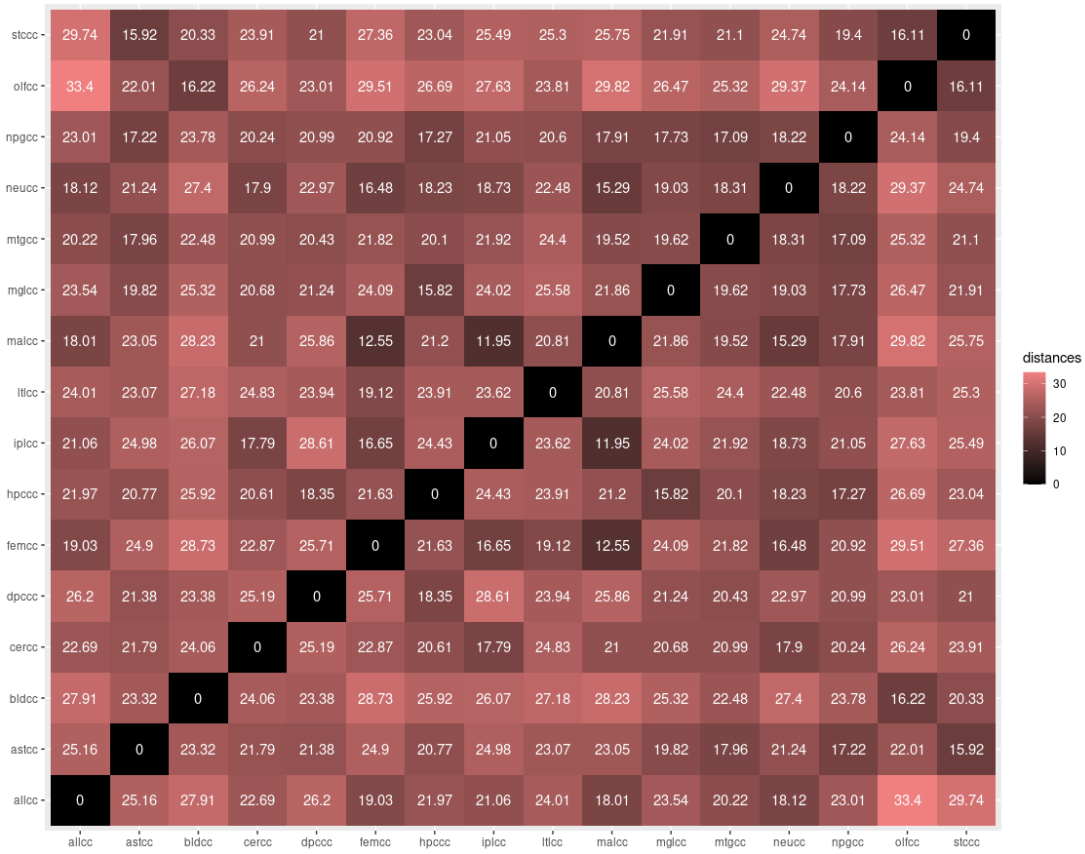

Figure 135 - Distance matrix of all sample groups, built upon CC

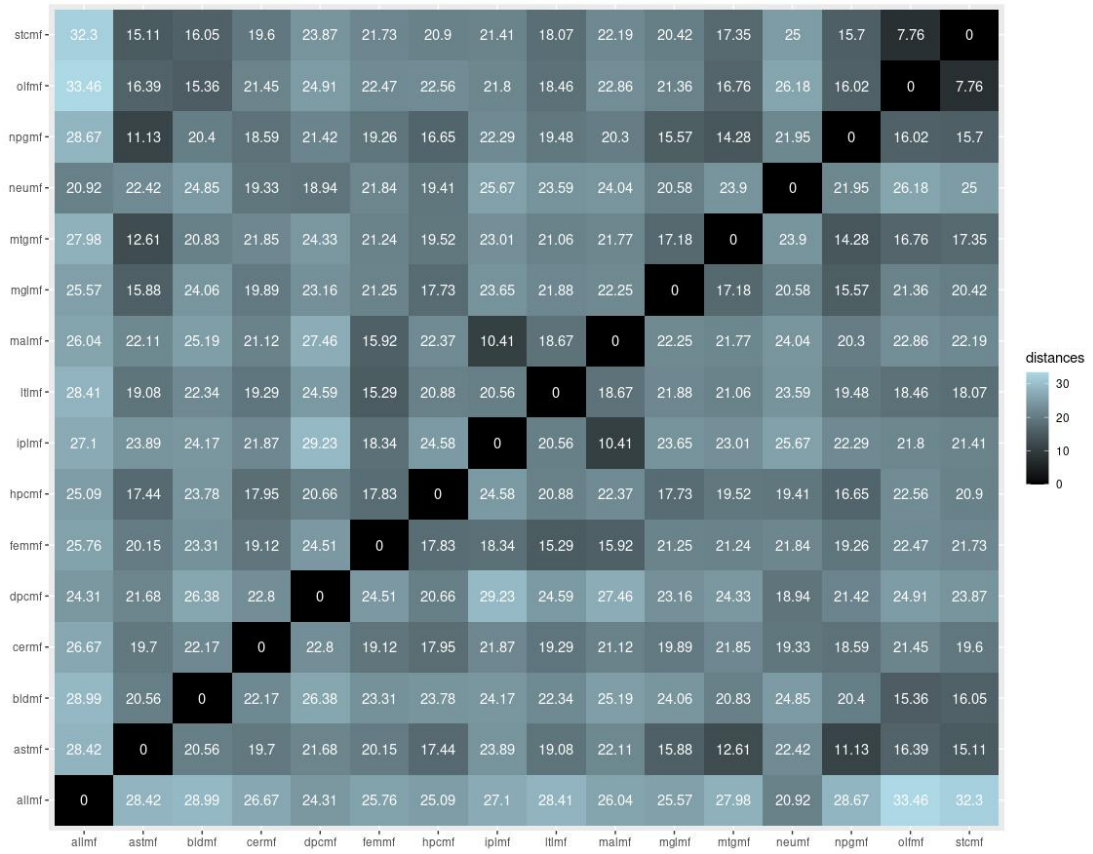

Figure 136 – Distance matrix of all sample groups, built upon MF

9.3. Topologies of All Samples and Comparisons

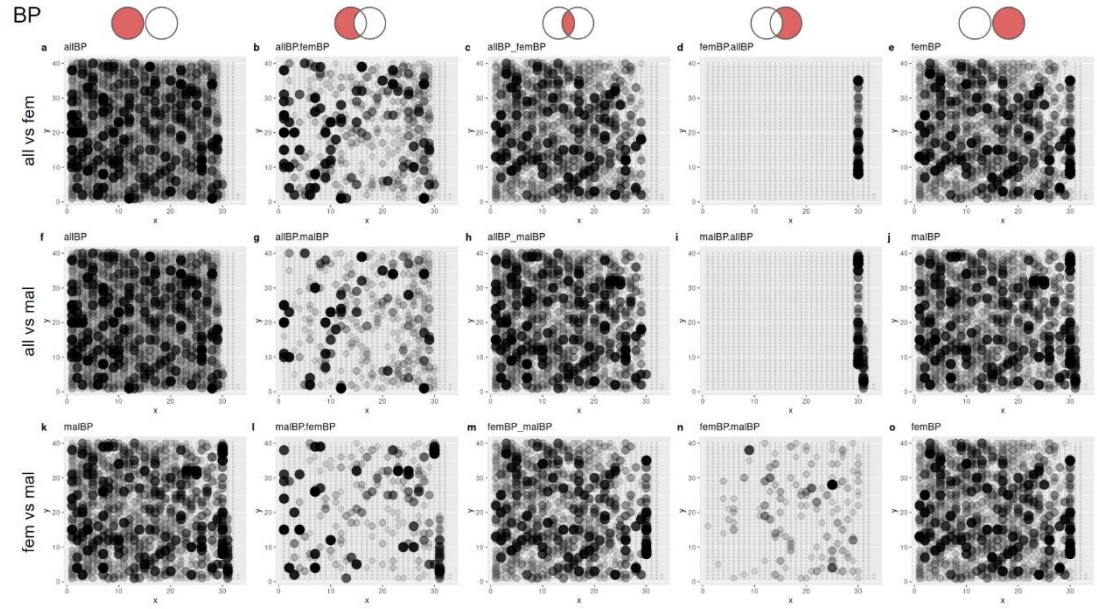

Figure 137 – BP topologies and comparisons of tissues grouped by sex

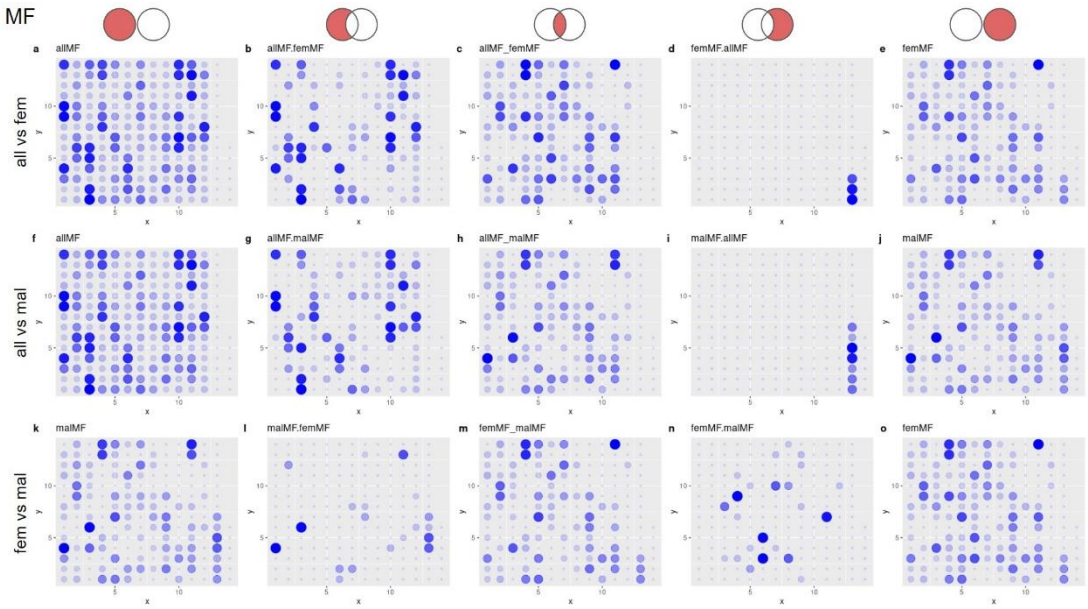

Figure 138 - MF topologies and comparisons of tissues grouped by sex

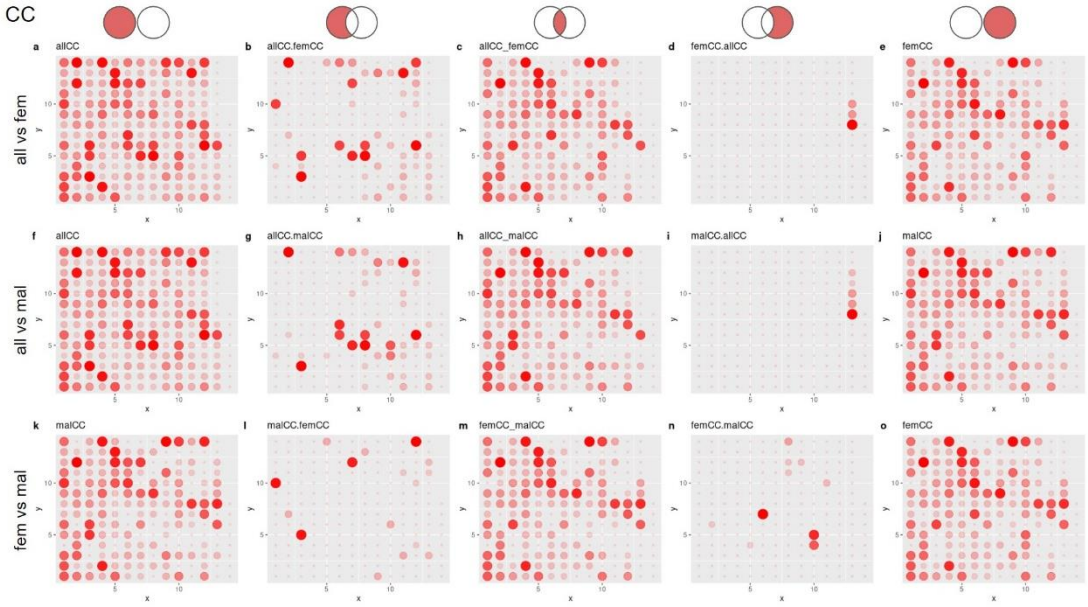

Figure 139 - CC topologies and comparisons of tissues grouped by sex

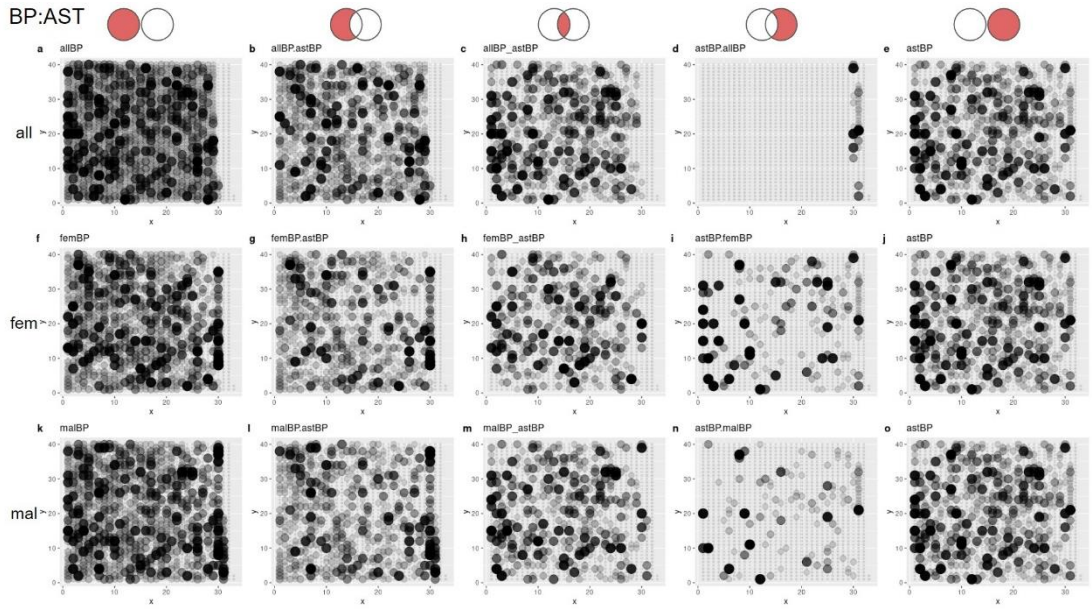

Figure 140 - BP topologies and comparisons of astrocytes vs. tissues grouped by sex

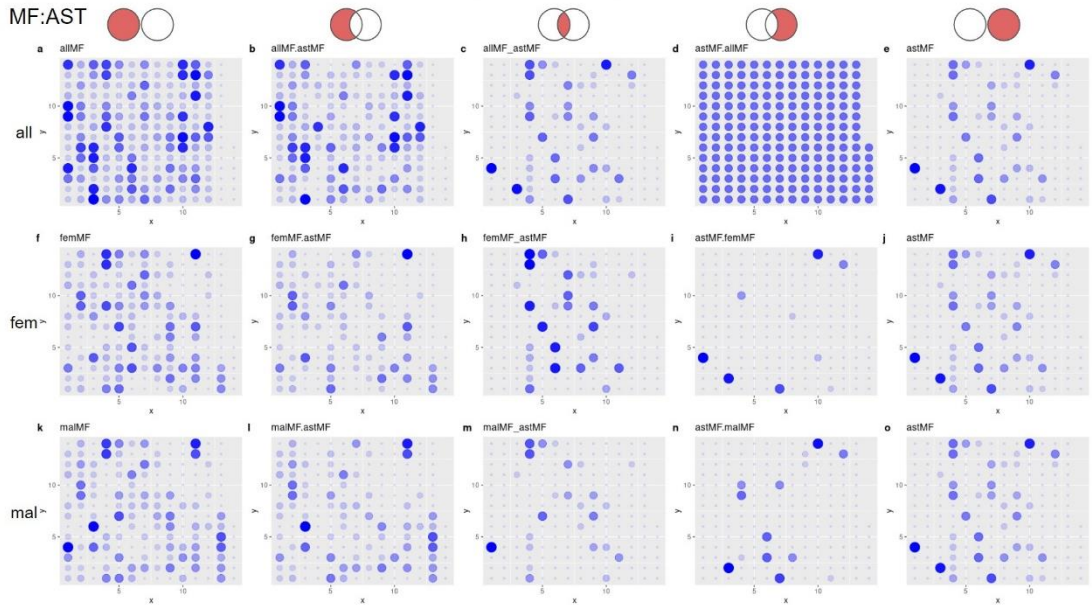

Figure 141 - MF topologies and comparisons of astrocytes vs. tissues grouped by sex

Figure 142 - CC topologies and comparisons of astrocytes vs. tissues grouped by sex

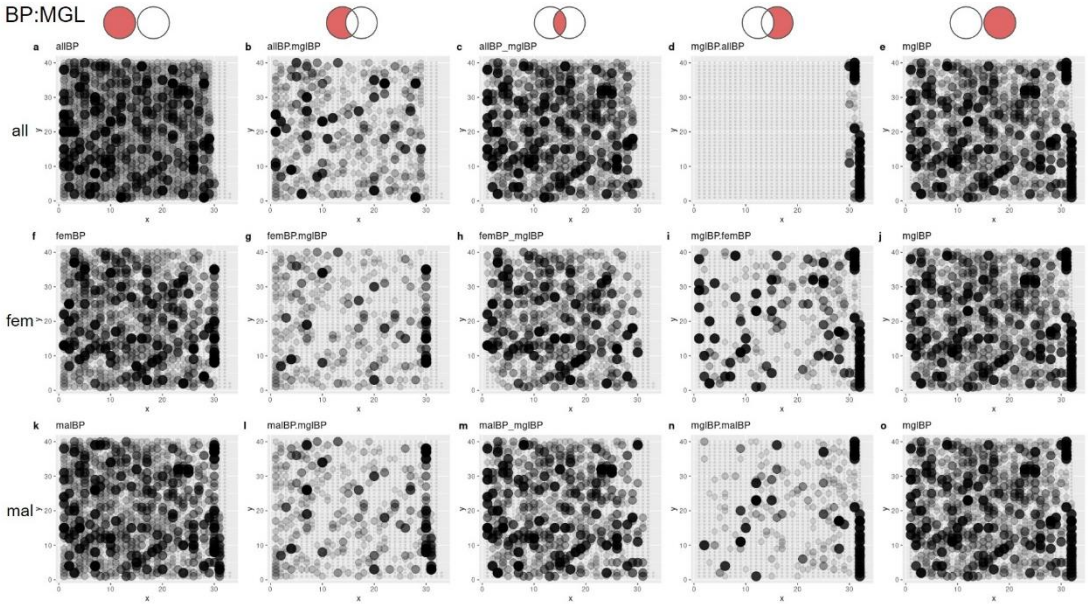

Figure 143 - BP topologies and comparisons of microglia vs. tissues grouped by sex

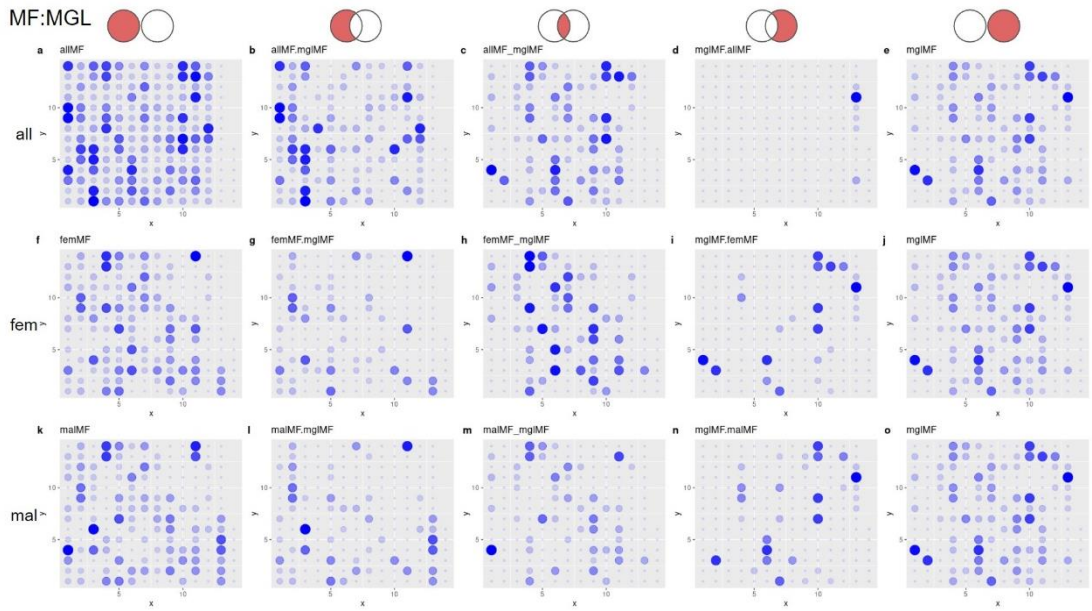

Figure 144 - MF topologies and comparisons of microglia vs. tissues grouped by sex

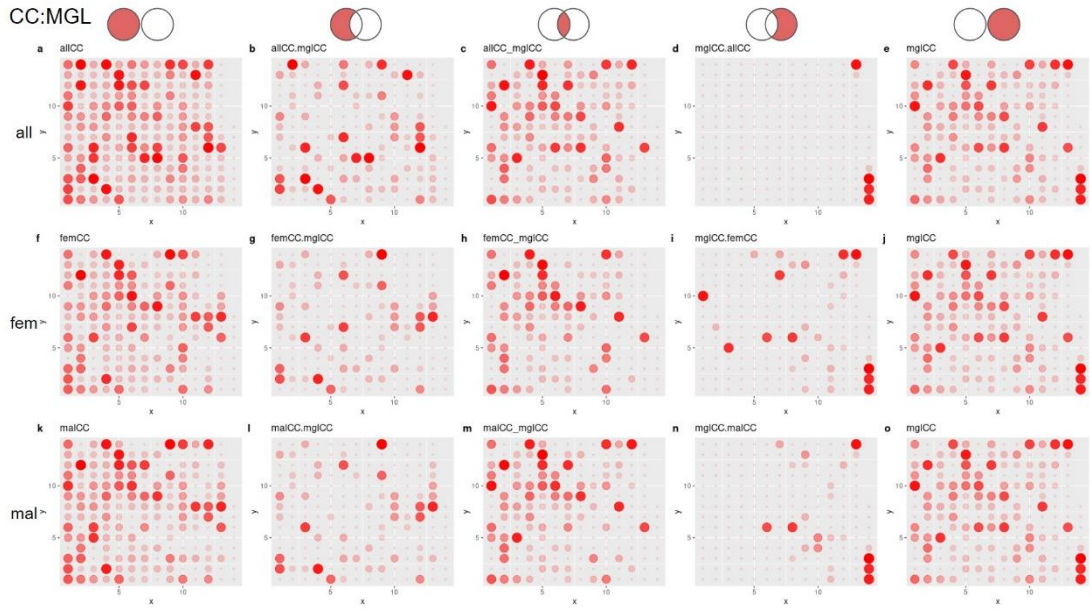

Figure 145 - CC topologies and comparisons of microglia vs. tissues grouped by sex

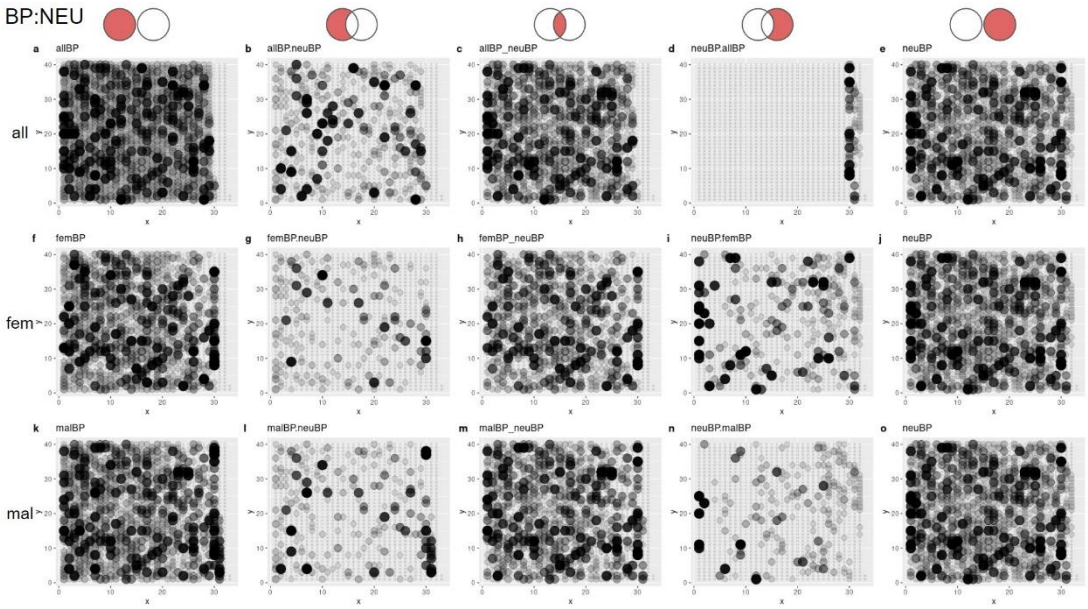

Figure 146 - BP topologies and comparisons of neurons vs. tissues grouped by sex

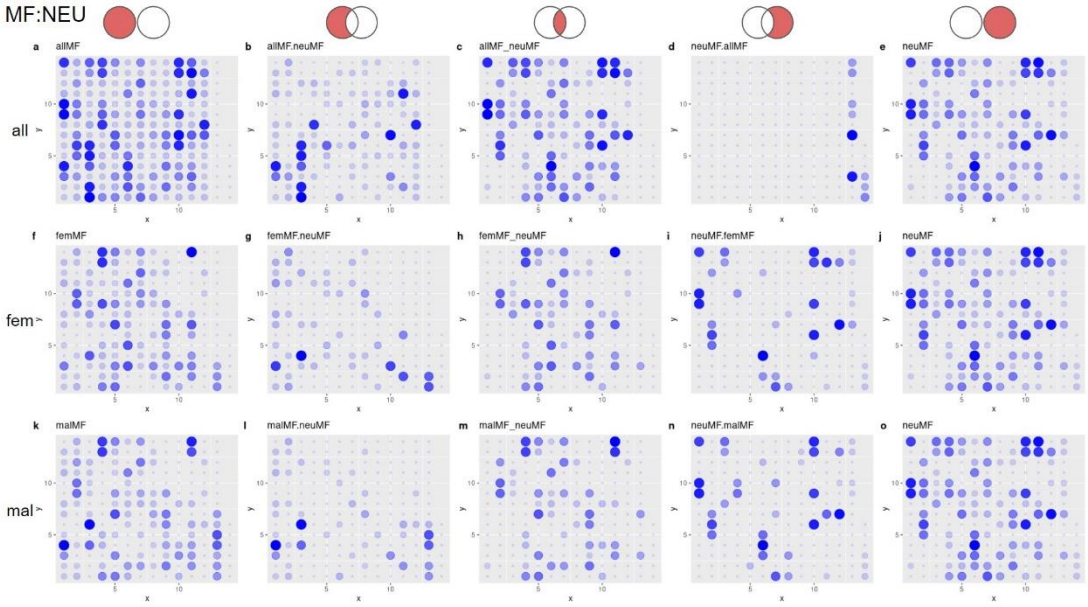

Figure 147 - MF topologies and comparisons of neurons vs. tissues grouped by sex

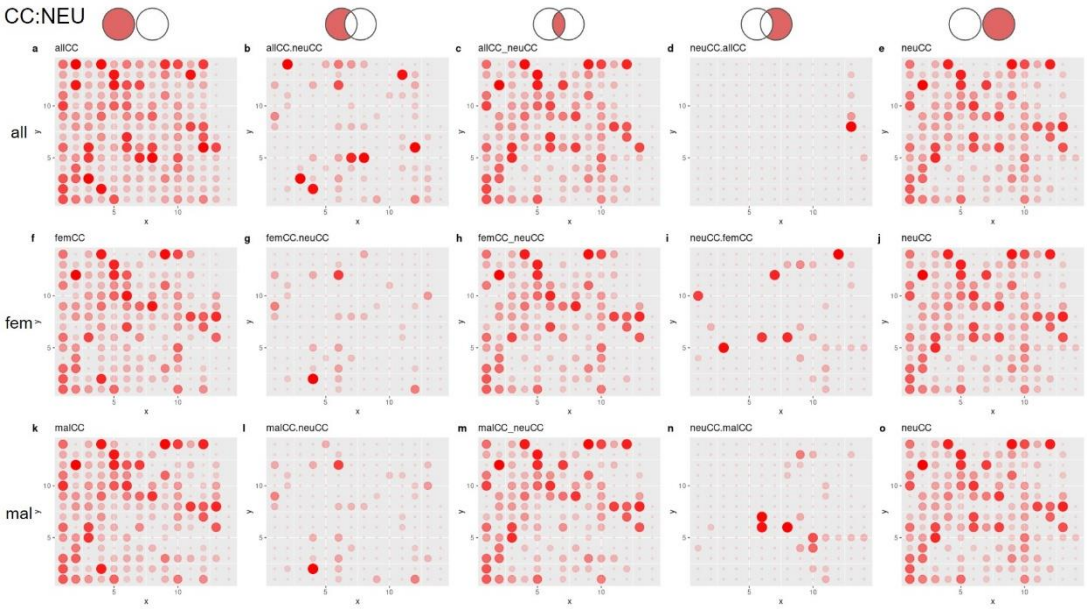

Figure 148 - CC topologies and comparisons of neurons vs. tissues grouped by sex

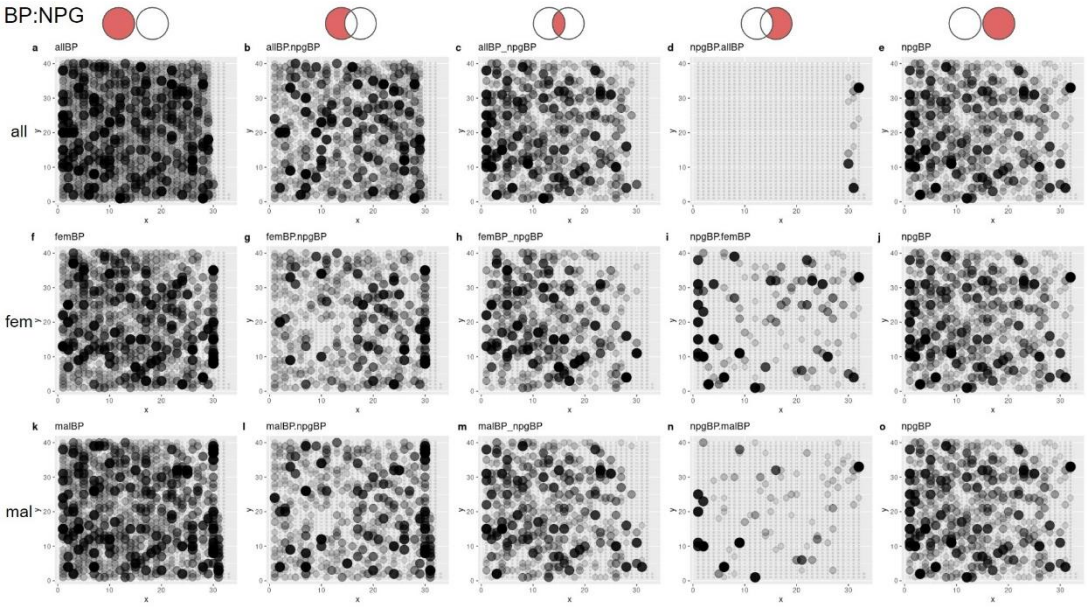

Figure 149 - BP topologies and comparisons of neural progenitors vs. tissues grouped by sex

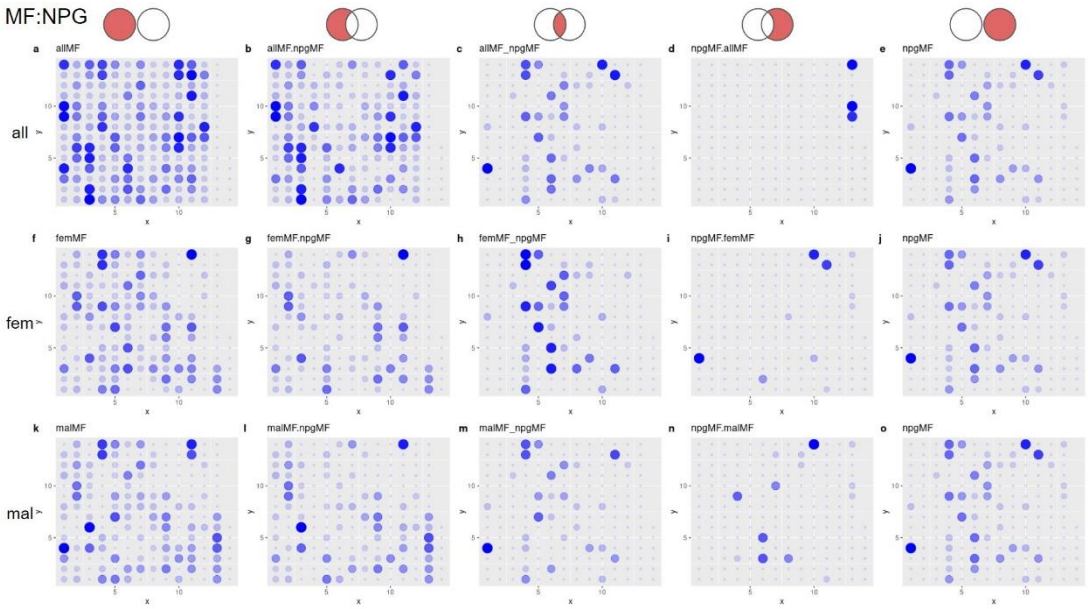

Figure 150 - MF topologies and comparisons of neural progenitors vs. tissues grouped by sex

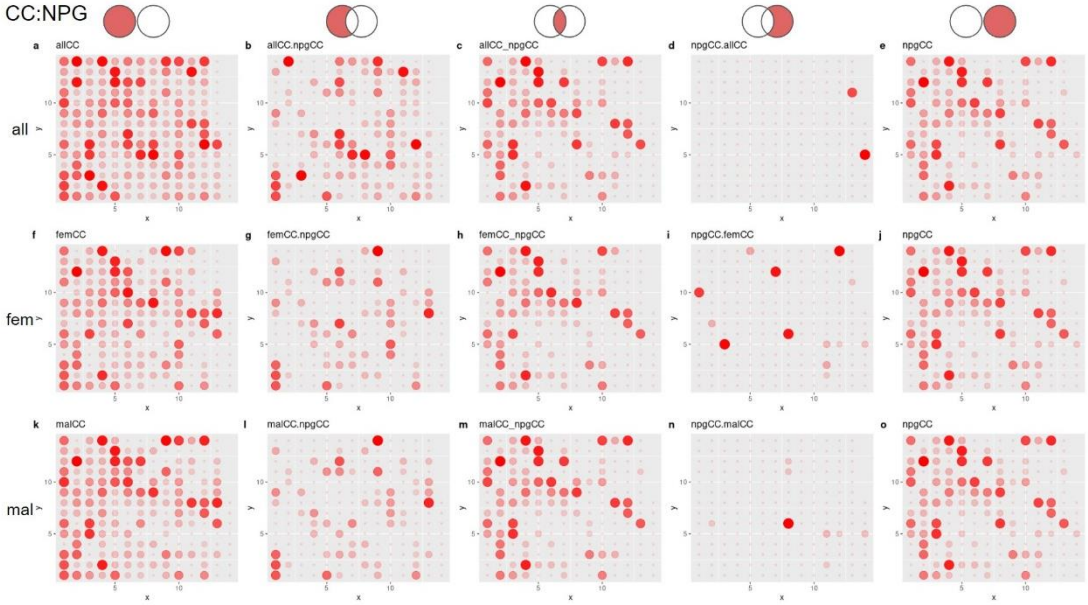

Figure 151 - CC topologies and comparisons of neural progenitors vs. tissues grouped by sex

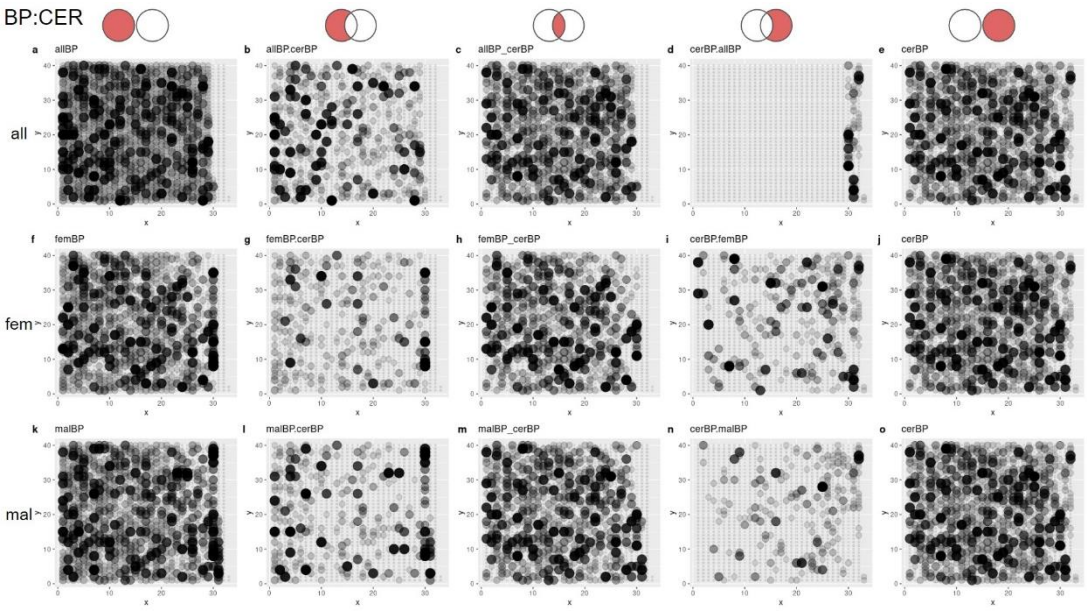

Figure 152 - BP topologies and comparisons of cerebral organoids vs. tissues grouped by sex

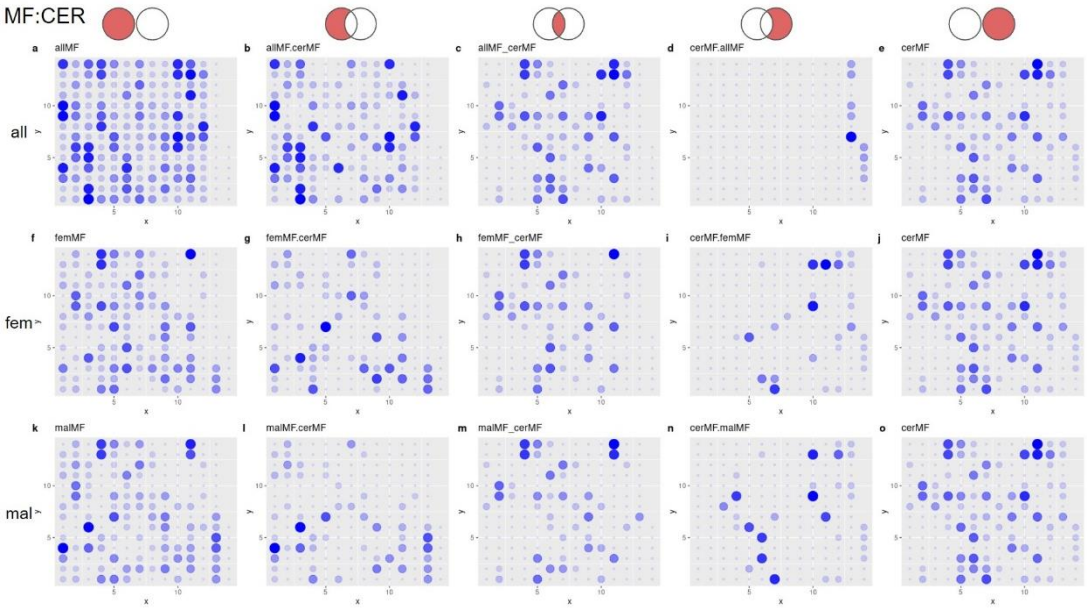

Figure 153 - MF topologies and comparisons of cerebral organoids vs. tissues grouped by sex

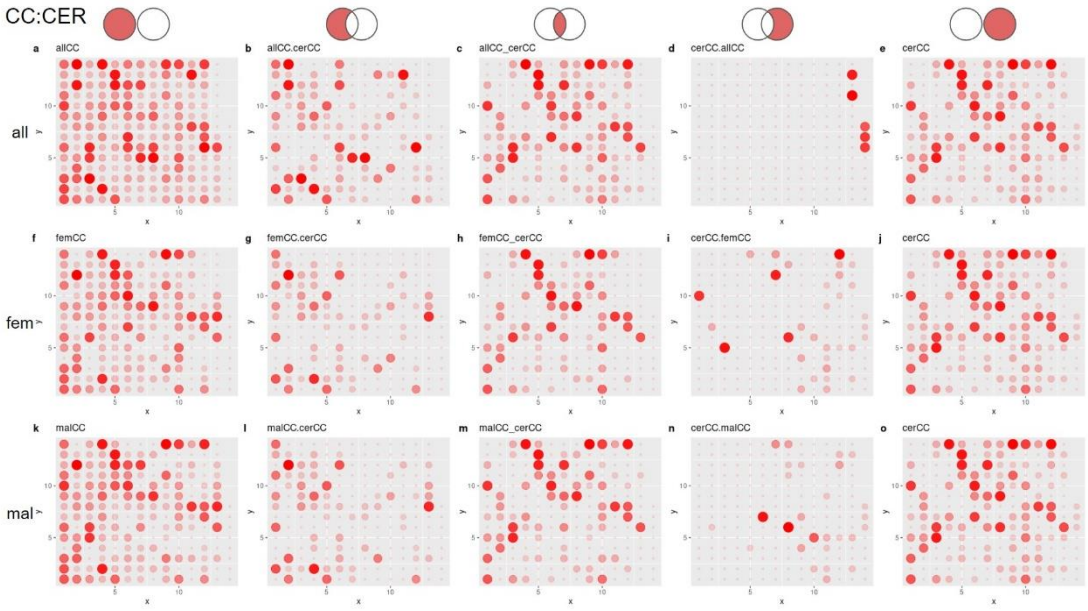

Figure 154 - CC topologies and comparisons of cerebral organoids vs. tissues grouped by sex

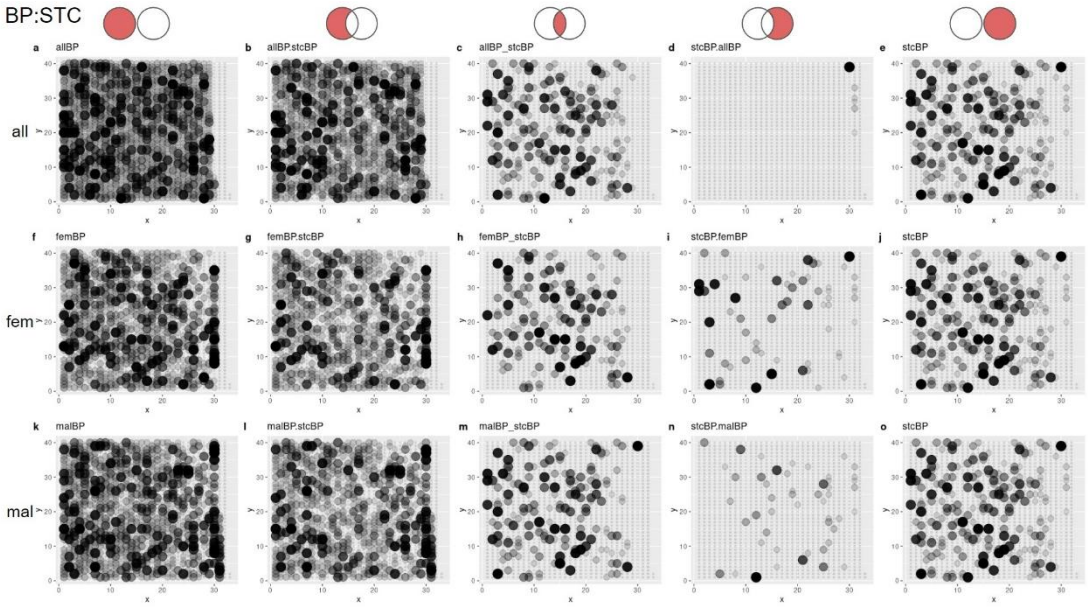

Figure 155 - BP topologies and comparisons of iPSCs vs. tissues grouped by sex

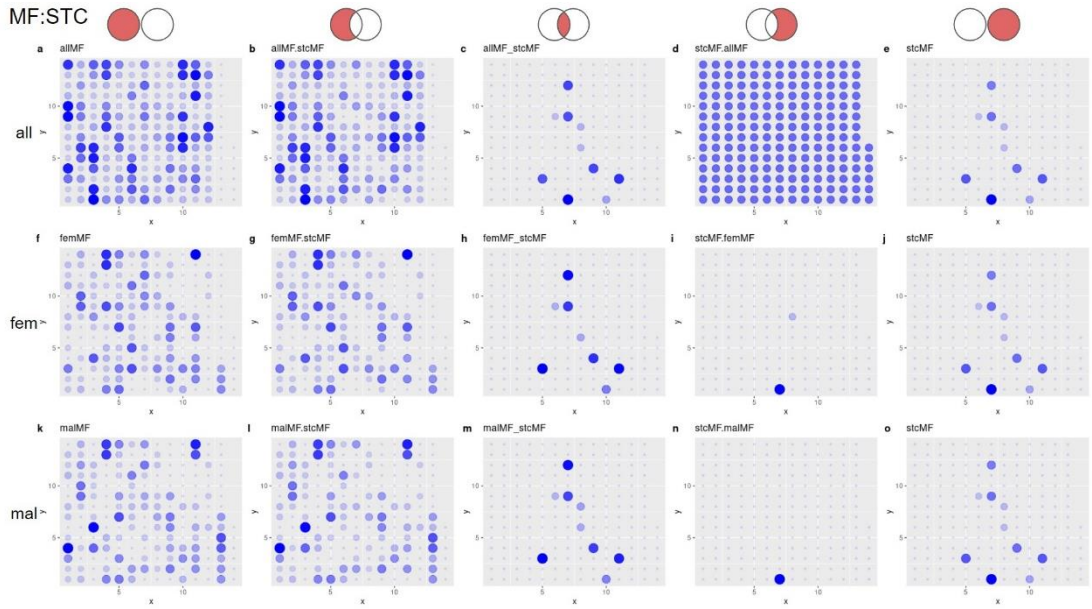

Figure 156 - MF topologies and comparisons of iPSCs vs. tissues grouped by sex

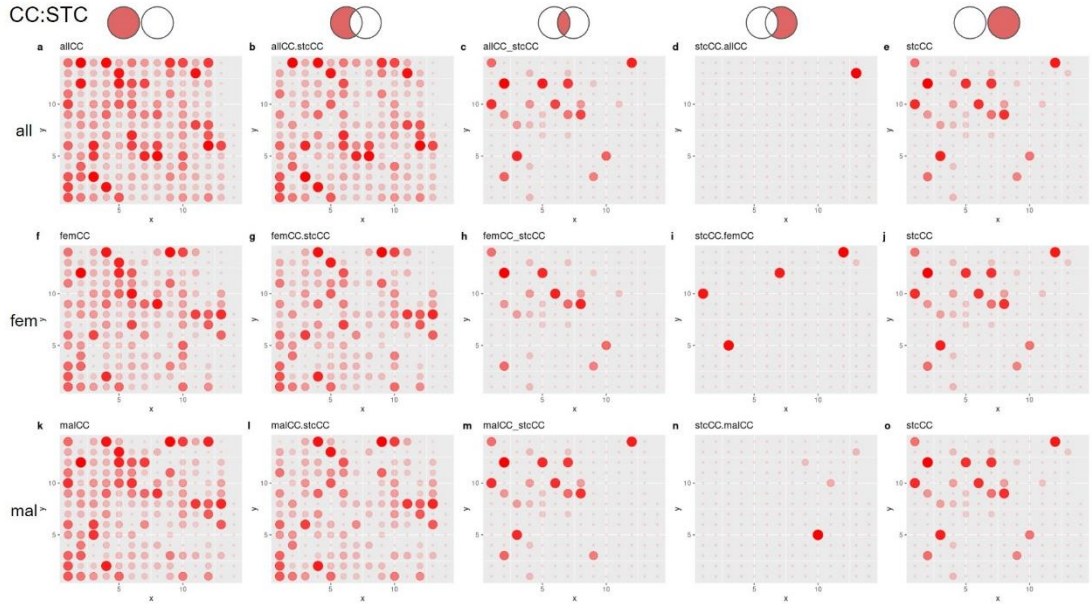

Figure 157 - CC topologies and comparisons of iPSCs vs. tissues grouped by sex

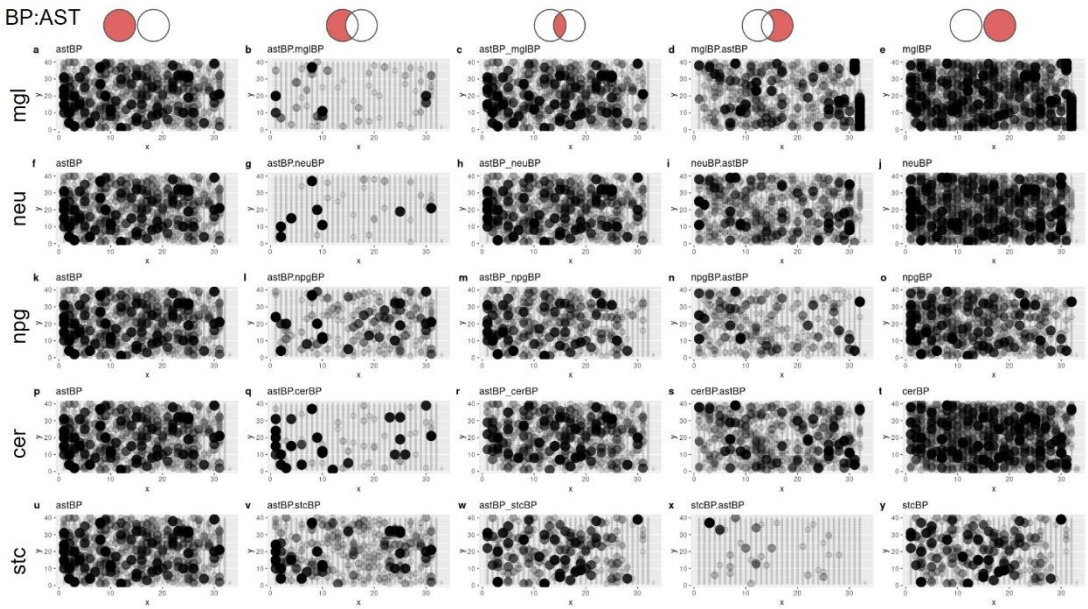

Figure 158 - BP topologies and comparisons of astrocytes vs. other cells

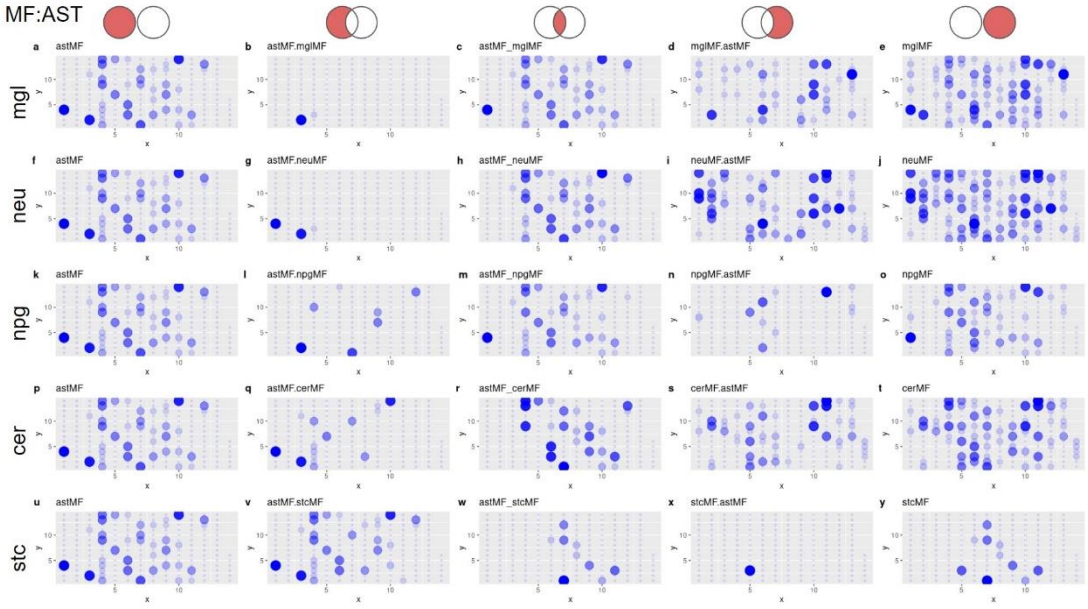

Figure 159 - MF topologies and comparisons of astrocytes vs. other cells

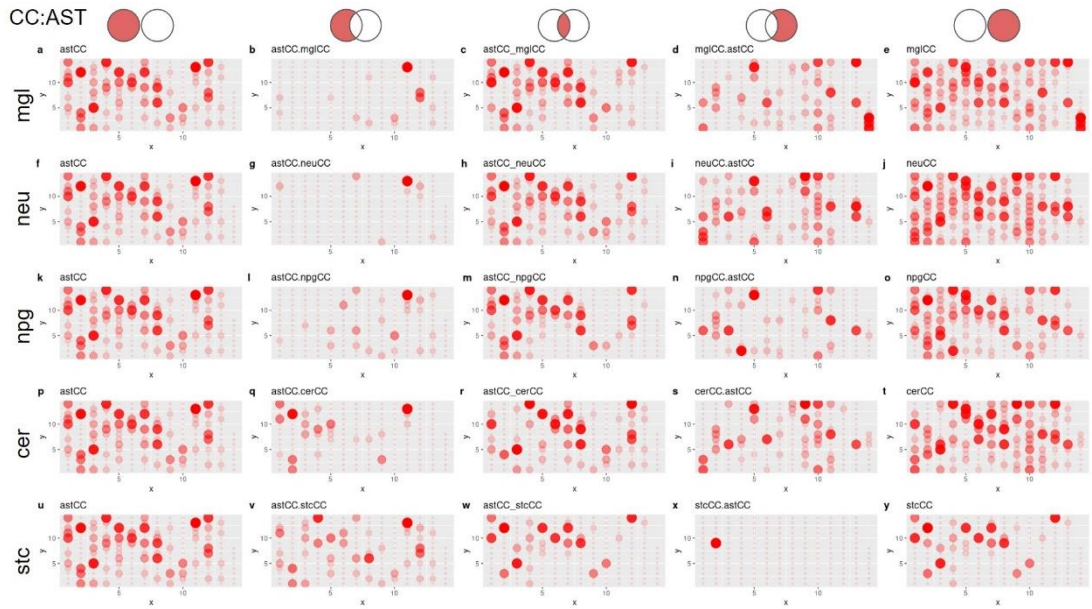

Figure 160 - CC topologies and comparisons of astrocytes vs. other cells

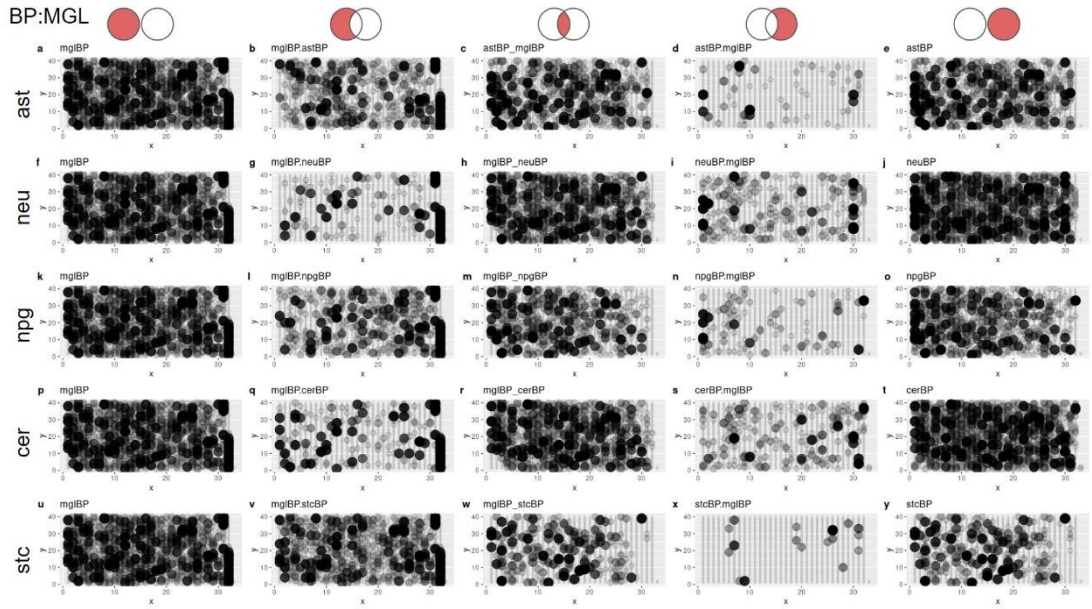

Figure 161 - BP topologies and comparisons of microglia vs. other cells

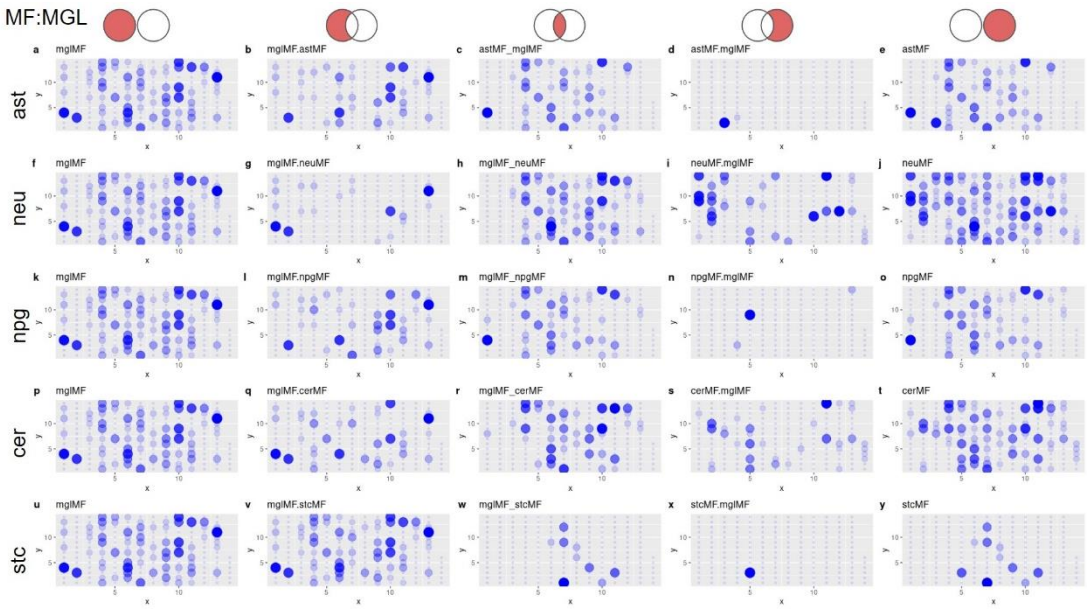

Figure 162 - MF topologies and comparisons of microglia vs. other cells

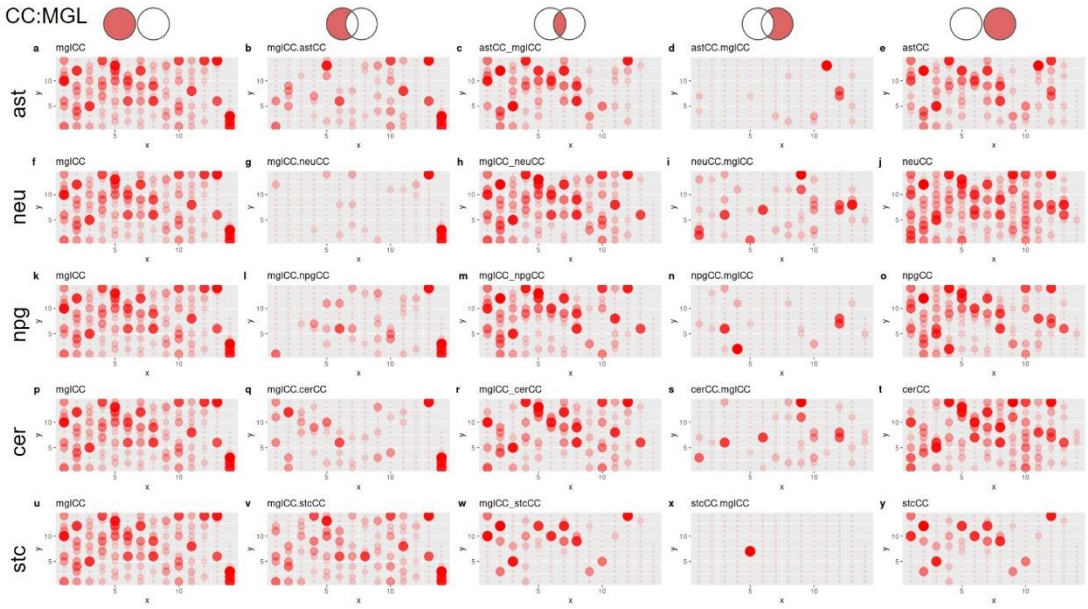

Figure 163 - CC topologies and comparisons of microglia vs. other cells

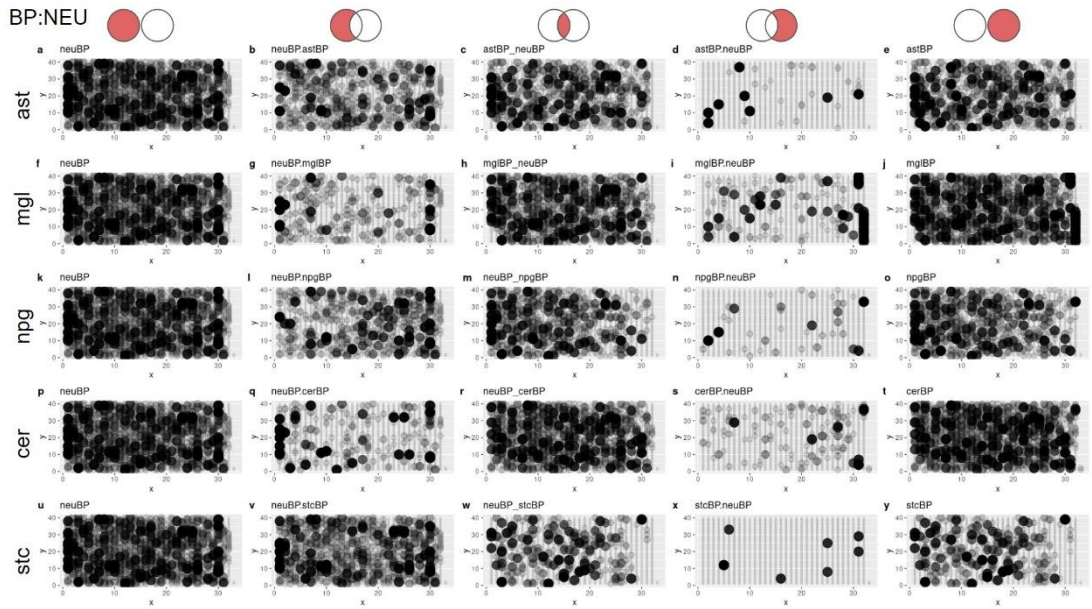

Figure 164 - BP topologies and comparisons of neurons vs. other cells

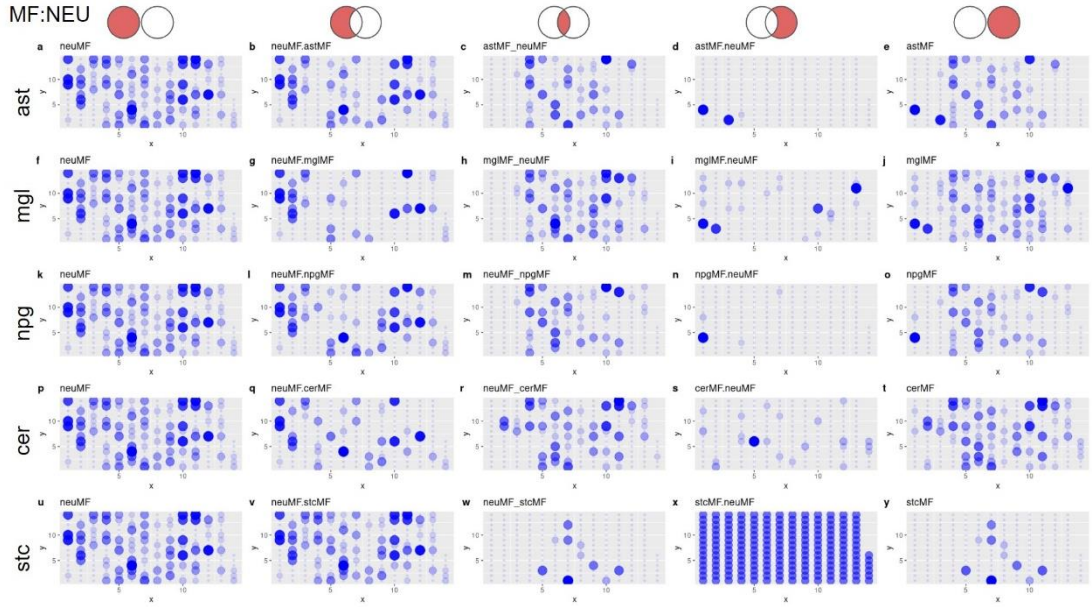

Figure 165 - MF topologies and comparisons of neurons vs. other cells

Figure 166 - CC topologies and comparisons of neurons vs. other cells

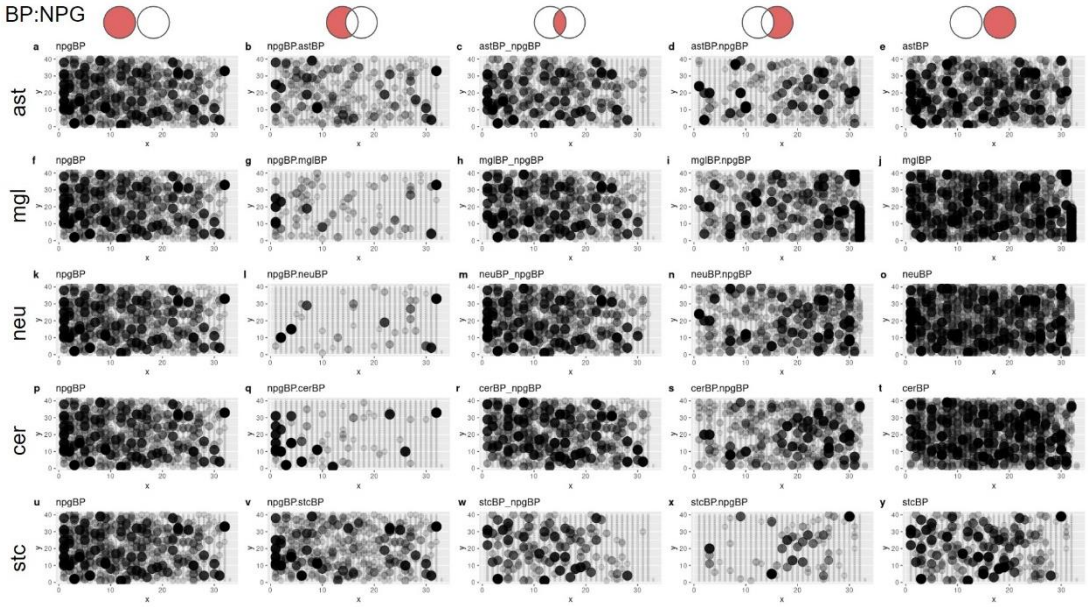

Figure 167 - BP topologies and comparisons of neural progenitors vs. other cells

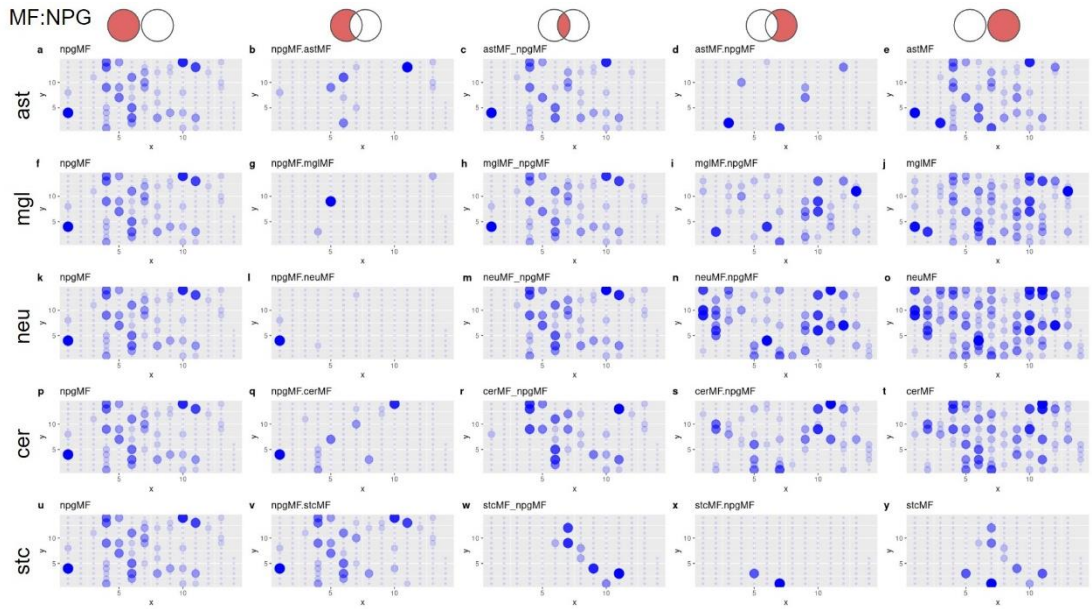

Figure 168 - MF topologies and comparisons of neural progenitors vs. other cells

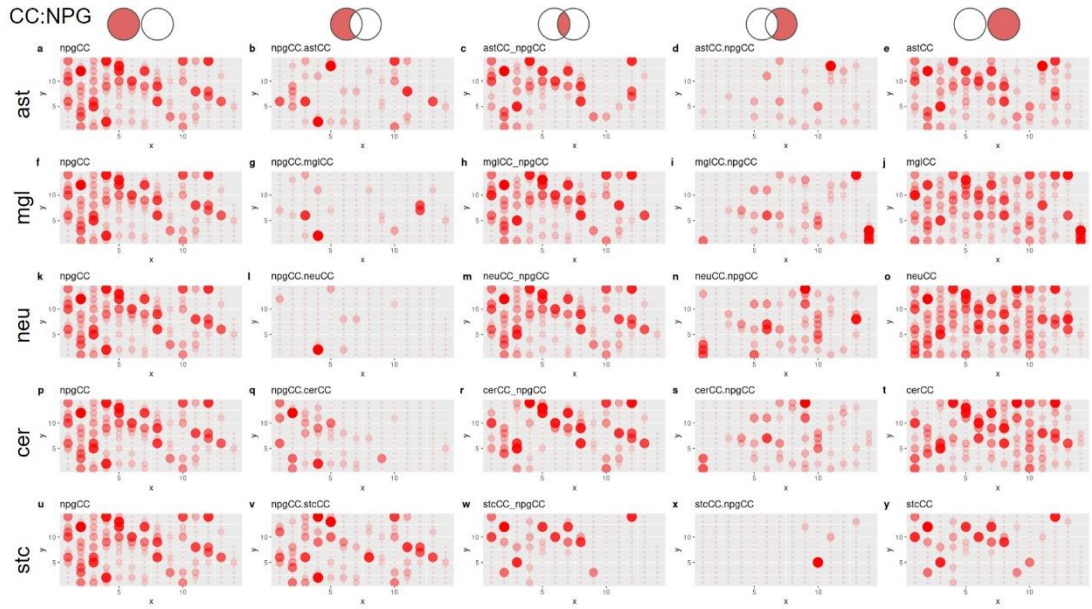

Figure 169 - CC topologies and comparisons of neural progenitors vs. other cells

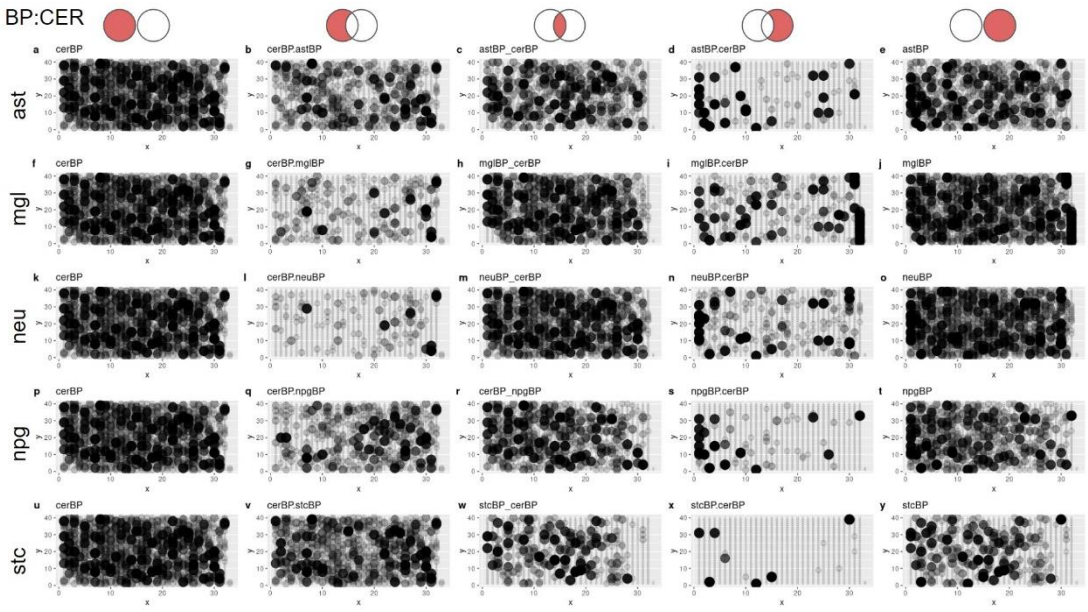

Figure 170 - BP topologies and comparisons of cerebral organoids vs. other cells

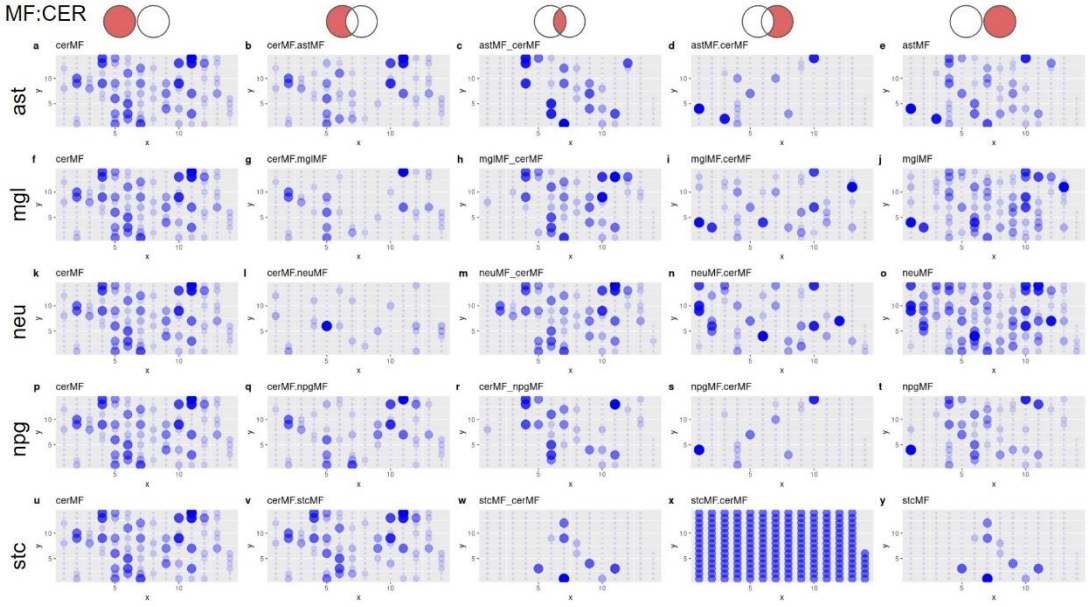

Figure 171 - MF topologies and comparisons of cerebral organoids vs. other cells

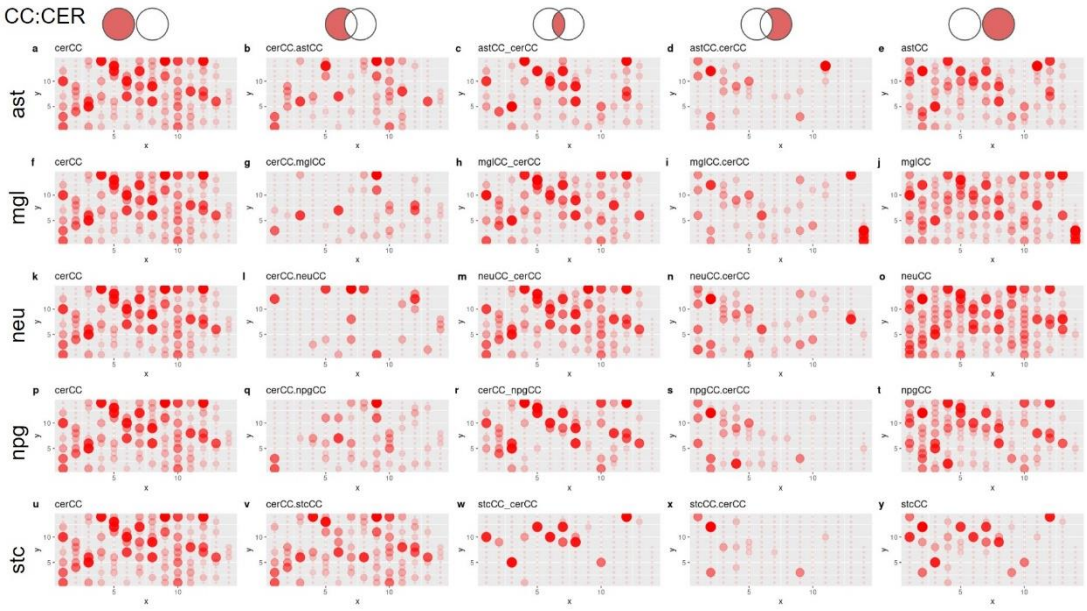

Figure 172 - CC topologies and comparisons of cerebral organoids vs. other cells

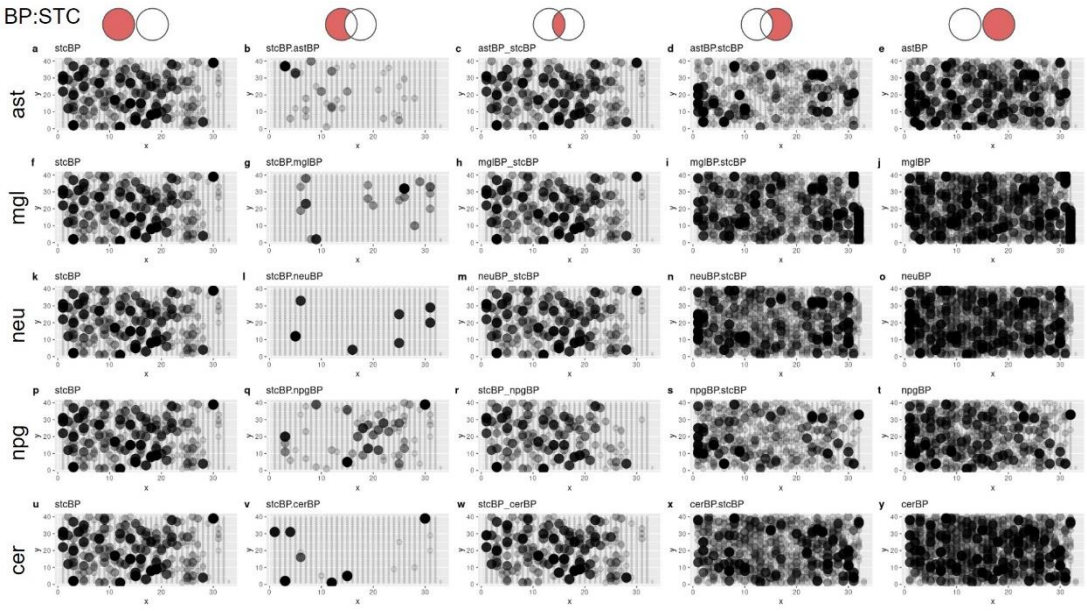

Figure 173 - BP topologies and comparisons of iPSCs vs. other cells

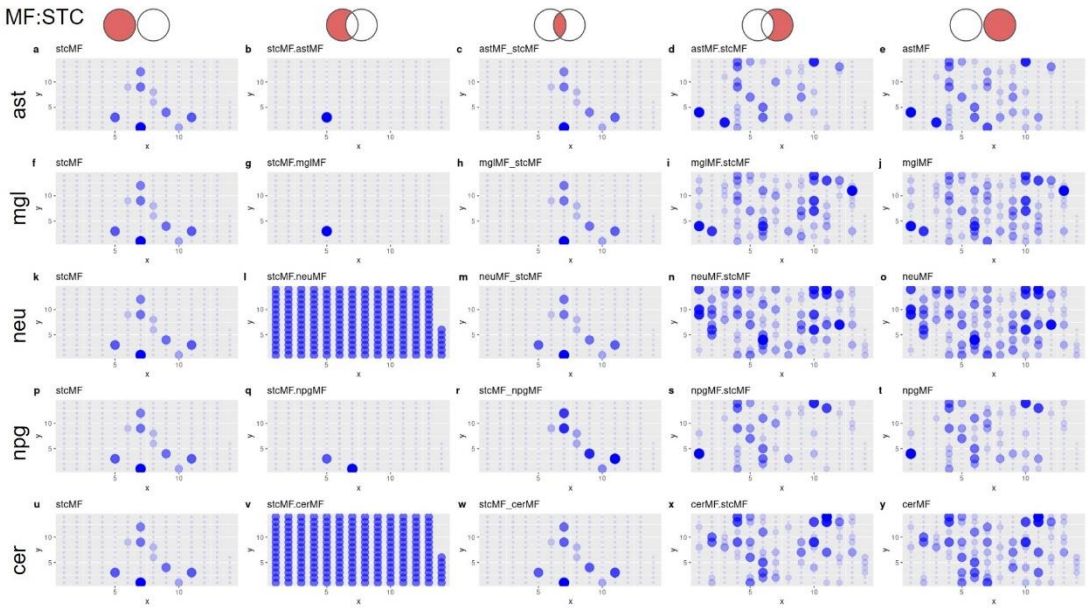

Figure 174 - MF topologies and comparisons of iPSCs vs. other cells

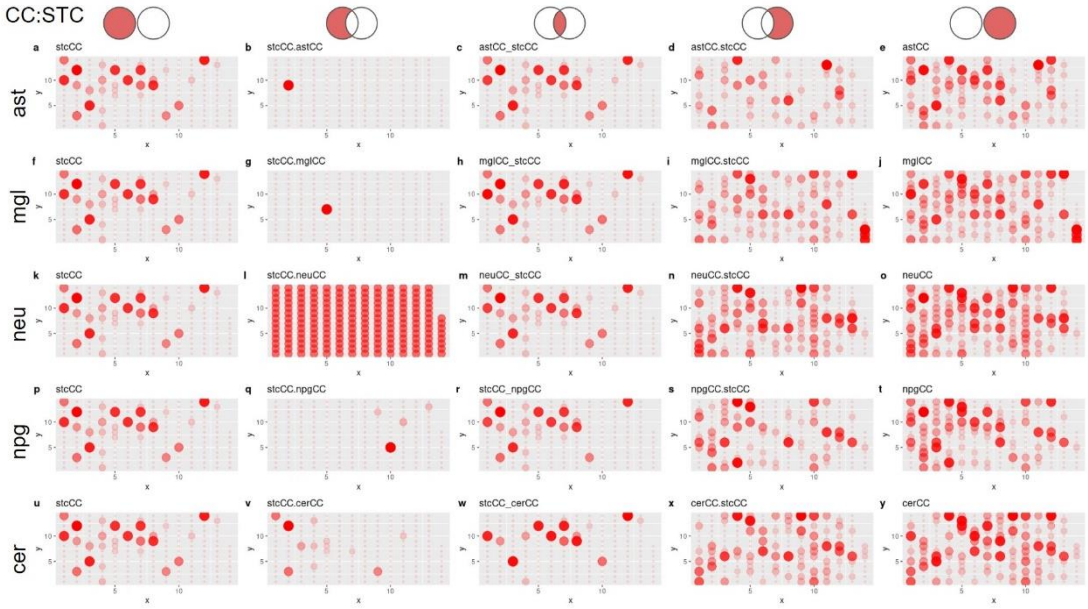

Figure 175 - CC topologies and comparisons of iPSCs vs. other cells