Imperial College London
Department of Brain Sciences

Evaluating Disorders of Cognition Using Integrative Genomics

Rahel Feleke

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)
June 2023

Copyright Declaration

Statement of Originality

I hereby declare that the work presented in this document is original and independently conducted by me and my collaborators, as mentioned in the relevant sections. All sources of information and references used in this work have been properly acknowledged and cited. The ideas, concepts, and interpretations presented in this document are my own, and two papers have previously been submitted to two journals, Acta Neuropathologica and Brain. I take full responsibility for the authenticity and originality of this work.

The two publications mentioned in this work are as follows:

Feleke, R., Jazayeri, D., Abouzeid, M., Powell, K.L., Srivastava, P.K., O'Brien, T.J., Jones, N.C. and Johnson, M.R., 2022. Integrative genomics reveals pathogenic mediator of valproate-induced neurodevelopmental disability. Brain, 145(11), pp.3832-3842.

Feleke, R., Reynolds, R.H., Smith, A.M., Tilley, B., Taliun, S.A.G., Hardy, J., Matthews, P.M., Gentleman, S., Owen, D.R., Johnson, M.R., Srivastava, P.K. and Mina, R., 2021. Cross-platform transcriptional profiling identifies common and distinct molecular pathologies in Lewy body diseases. Acta Neuropathologica, 142, pp.449-474.

Acknowledgement

I would like to begin by thanking my Lord and Saviour, Jesus Christ. Through His providence, I have been able to overcome challenges and achieve success in my research endeavours. His steadfast presence has been a constant source of strength and comfort, and I am humbled by His grace and blessings.

I wish to express my sincerest thanks to the individuals who graciously donated their brain tissue, as their contribution has been instrumental in completing this thesis. Their generous contributions have enabled us to make progress in our understanding of neurological diseases. It is my hope that this research will honour their generosity by advancing our knowledge and ultimately leading to treatments for these devastating conditions.

I am incredibly fortunate to have had the support and mentorship of two great scientists, who have contributed significantly to the trajectory of my career. I would like to extend my heartfelt appreciation to my primary supervisor, Professor Michael Johnson, for his unwavering patience and steadfast guidance throughout my PhD journey. His expertise in systems biology, as well as his commitment to my development and improvement, have been important in moulding my research skills and assisting me in becoming a better research scientist. I am particularly grateful to Dr Prashant Srivastava, who gave me invaluable assistance and direction in assessing and analysing complex data. His advice and willingness to share his knowledge have been instrumental in the success of my research.

I am grateful to have been a part of a collaborative and dynamic research community that provided me with a learning and growing environment. I would like to thank my colleagues Jiahui Ji, Maya Abouzeid, Alexander Haglund, Jeong Hyun Ko, Ann C. Babtie, Liv Wiemann and Leonardo Bottolo for their support and friendship. Their constructive feedback, insightful discussions, and collaborative efforts have significantly contributed to my professional development and growth.

I shall be forever grateful to my dear family and friends, whose unwavering encouragement and support have been a constant source of comfort and strength to me. Their trust in my abilities

has been the foundation of my success. To my dear parents and siblings, I cannot express my gratitude enough for their unconditional love. They have been my guiding lights, forever inspiring and motivating me to be my best self.

Finally, I would like to acknowledge the generous funding support of UKRI MRC for making my research possible. I would also like to extend my gratitude to Imperial College London for providing me with a stimulating and supportive research environment. The facilities, resources, and expertise available at Imperial College have been instrumental in facilitating my research and professional development.

Abstract

Integrative genomics embodies a collaborative approach that brings together various disciplines, merging genomic and computational methodologies to achieve a comprehensive understanding of complex biological systems. By integrating and analysing various omics datasets, researchers can uncover novel insights into biological processes, disease mechanisms, and molecular interactions.

This thesis applies integrative genomic approaches to study the biological mechanisms of cognitive disorders resulting from neurodegenerative and neurodevelopmental impairments. It includes three studies that utilise transcriptomic profiling to identify and assess these mechanisms. The final discussion chapter summarises the main findings, including shared and distinct mechanisms across different cognitive disorders, addressing technical limitations, and outlining future research directions.

The first study focuses on Lewy body diseases, utilising single-nucleus transcriptomics. Differential expression analysis reveals widespread dysregulation in neurons and glial cell types, with similar gene expression profiles observed in Parkinson's Disease Dementia (PDD) and Dementia with Lewy Bodies (DLB), while Parkinson's Disease (PD) shows distinct transcriptional profiles. Heritability enrichment analysis highlights a genetic association between glial cell dysregulation and PD age of onset. Additionally, a unique population of neurons associated with DLB, resembling medium spiny neurons, is identified.

The second study investigates early-stage abnormal tau species-related gene expression changes in Alzheimer's Disease (AD). Analysis of gene expression at the single-cell level demonstrates distinct patterns between Tau-proximity Ligation Assay (tauPLA) positive and tauPLA negative brain tissues. Dysregulation of genes in various cell-types in tauPLA positive samples is observed in the absence of neurofibrillary tangles. Reactive astrocyte activation, even in the absence of neurofibrillary tangles, is also reported for the first time in a transcriptomic-based study, suggesting their role in AD progression.

The third study examines the neurodevelopmental outcomes of valproate exposure. Transcriptomic analysis reveals significant gene expression changes in the brains of gestationally exposed pups, affecting synaptic function, neurodevelopment, and genes associated with schizophrenia,

bipolar disorder, and IQ heritability. Differential splicing analysis suggests enduring effects on brain function through epigenetic encoding.

Convergent and divergent mechanisms underlying various neurodegenerative disorders are identified across the studies, including pathways related to axonal degeneration, mRNA splicing, synaptic organisation, autophagy, neuron death, phosphorylation, memory, mitochondrial function, and vesicle-mediated transport regulation.

This thesis contributes to understanding the biological basis of cognitive disorders through integrative genomic approaches, providing insights into shared and distinct mechanisms across different disorders. The findings have implications for the development of therapeutic interventions and underscore the importance of rigorous experimental design in transcriptomic investigations. Future research directions are outlined, aiming to further unravel the complex molecular mechanisms underlying cognitive disorders.

List of Abbreviations

Aβ amyloid beta. 66

A1 Reactive astrocytes. 94

AD Alzheimer's Disease. v, 1, 4, 19, 65, 66, 134

ADHD Attention-deficit/hyperactivity disorder. 17, 106, 116

AEDs Antiepileptic drugs. 99, 100

All-E All Exposed. 112, 114

All-N All Non-Exposed. 112, 114

AMPA α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid. 100

APOE Apolipoprotein E. 29

APOE4 Apolipoprotein E 4. 29

APP Amyloid Precursor Protein. 66, 92

ARIDIB AT-rich interactive domain-containing protein 1B. 99

ASD Autism Spectrum Disorder. 17, 99, 101, 106, 116, 118, 124, 125, 133, 134

ASH1L ASH1 like histone lysine methyltransferase. 99

ATP13A2 ATPase Cation Transporting 13A2. 28, 61

ATP2B2 ATPase Plasma Membrane Ca2+ Transporting 2. 124

BD Bipolar Disorder. 106, 116, 126

BH Benjamin and Hochberg. 37, 77, 106

BIN1 Bridging Integrator 1. 29

BMI body mass index. 30

C-section Caesarean section. 104

CACNA1A calcium voltage-gated channel subunit alpha1 A. 45

CACNA1H Calcium channel, voltage-dependent, T type, alpha 1H subunit. 100

CACNG8 calcium voltage-gated channel auxiliary subunit gamma 8. 45

CAMATA1 Calmodulin Binding Transcription Activator 1. 118

CAMK2B Calcium/Calmodulin De- pendent Protein Kinase II Beta. 45

CAMK2D Calcium/Calmodulin Dependent Protein Kinase II Delta. 45

CAMK4 Calcium/Calmodulin Dependent Protein Kinase IV. 124

CCA canonical correlation analysis. 9

CDG Cross Disorders Group. 106, 116, 126

CDKL5 Cyclin- Dependent Dinase-Like 5. 56

cDNA complementary DNA. 4, 13, 34

CDR Cellular Detection Rate. 13, 37, 43

CHCHD2 coiled-coil helix coiled-coil helix domain 2. 28

CHD1 chromodomain helicase DNA binding protein 1. 98

CHD8 chromodomain helicase DNA binding protein 8. 98

CMC Combined Multivariate and Collapsing Rare Variant Analysis. 14

CNS Central nervous systems. 17, 94, 99

CNVs copy number variations. 98

Conos Clustering On Network Of Samples. 9

CPM counts per million. 10

CSF cerebrospinal fluid. 66

DDX3Y DEAD-Box Helicase 3 Y-Linked. 105

DE Differential expression. 43

DEA Differential Expression Analysis. 13, 36, 39, 83

DEGs Differentially Expressed Genes. 41, 83, 105, 108, 112, 114, 134

DLB Dementia with Lewy Bodies. v, 19, 20, 24, 33, 36, 37, 40, 134

DLG2 discs large MAGUK scaffold protein 2. 45

DNA Deoxyribonucleic acid. 3, 4

DNAJC13 homolog subfamily C member 13. 28

DNAJC6 DNAJ/HSP40 homolog subfamily C member 6. 28, 60

DPX Distyrene, a Plasticizer, and Xylene. 73

DSCAM DS cell adhesion molecule. 98

DSM-5 Diagnostic and Statistical Manual of Mental. 17, 98

DUBs Deubiquitinating enzymes. 93

DWLS Dampened Weighted Least Squares. 108, 111

E-GAERS Exposed Genetic Absence Epilepsy Rats from Strasbourg. 105, 108, 112

E-NEC Exposed Non-Epileptic Control. 105, 108, 112

EIF4G1 Eukaryotic translation initiation factor 4 gamma 1. 28

EMBL-EBI The European Molecular Biology Laboratory-European Bioinformatics Institute. 107

EPI Epilepsy. 106, 116

eQTLs expression quantitative trait loci. 144

ERK extracellular signal-regulated kinase. 101

ERK1 Extracellular Signal-Regulated Kinase 1. 124

ERK2 Extracellular Signal-Regulated Kinase 2. 124

EWCE Expression Weighted Celltype Enrichment. 36, 38

FBX07 F-box only protein 7. 28

FDR false discovery rate. 37, 39

FET Fisher's exact test. 82

FFPE Formalin-Fixed Paraffin-Embedded. 72

FGS first-generation DNA sequencing technology. 3

FOXP2 Forkhead Box Protein P2. 56

FOXP4 Forkhead box P4. 118, 125

FPR false positive rate. 75

GABA Gamma-aminobutyric acid. 100, 101

GABAA gamma-aminobutyric acid A. 100

GABRB3 Gamma-Aminobutyric Acid Type A Receptor Subunit Beta3. 124

GAD2 glutamate decarboxylase 2. 118, 125

GAERS Genetic Absence Epilepsy Rats from Strasbourg. 102, 103, 110, 113, 116

GBA Glucosylceramidase Beta 1. 28, 29, 59

GGE Genetic Generalized Epilepsy. 100, 102

GO Gene Ontology. 13, 37, 77, 106

GSEA Gene-set enrichment analysis. 13

GWAS Genome-Wide Association Studies. 1, 14, 28, 31, 98, 116

H-MAGMA Hi-C-coupled MAGMA. 38, 78

H_2O_2 hydrogen peroxide. 73

HCN1 Hyperpolarization Activated Cyclic Nu- cleotide Gated Potassium Channel 1. 124

HDAC histone deacetylase. 101, 126

HES heritability enrichment score. 14

hESCs human embryonic stem cells. 99

HGP Human Genome Project. 3

hiPSCs human induced pluripotent stem cells. 99

HTRA2 HtrA serine peptidase 2. 28

ID intellectual disability. 17

IGF1R Insulinlike Growth Factor1 Receptor. 56

IQ intelligence quotient. 2, 101, 106, 116, 126

KCNJ6 G protein-activated inward rectifier potassium chan- nel 2. 45

KDM6B Lysine demethylase 6B. 99

KEGG Kyoto Encyclopedia of Genes and Genomes. 77

KMT2A Lysine [K]-specific Methyl Transferase 2A. 99

KNN K-Nearest Neighbour. 11

LBDs Lewy Body Diseases. 20, 24, 26, 32, 43, 128

LD Linkage Disequilibrium. 14

LDSC Linkage Disequilibrium Score Regression. 14, 15, 79, 88, 106, 107, 116

LIGER Linked Inference of Genomic Experimental Relationships. 9

LRRK2 eucine-rich repeat kinase2. 28, 29, 92

MAGMA Multi-marker Analysis of GenoMic Annotation. 14

MAPK1 Mitogen-Activated Protein Kinase 1. 56

MAPT icrotubule-associated protein tau. 29, 92

MAST Model-based Analysis of Single-cell Transcriptomics. 12, 36

MC1 monoclonal Receptor Polyclonal Antibody. 69

MDS Movement Disorder Societ. 33

MNNs mutual nearest neighbours. 8

MPPP 1-methyl-4-phenyl-4-propionoxypiperidine. 30

MPTP 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine. 30

MR Mendelian randomisation. 31

mRNA messenger RNA. 13, 24

MSN medium spiny neurons. 24, 60

mtRNA mitochondrial RNA. 140

N-GAERS Non-Exposed Genetic Absence Epilepsy Rats from Strasbourg. 105, 108, 112

N-NEC Non-Exposed Non-Epileptic Control. 105, 108, 112

NDDs Neurodevelopmental Disorders. 17, 98, 99

NDs Neurodegenerative Diseases. 19, 136

NEAD Neurodevelopmental Effects of Antiepileptic Drugs. 101

NEC Non-Epileptic Control. 102, 103, 113

NFTs neurofibrillary tangles. 66, 72, 73, 92, 95, 131

NGS next-generation sequencing. 3

NMDA N-metyl-D-aspartate. 101

NTs neuropil threads. 73

PARK7 Parkinsonism associated deglycase. 60, 92

PC principal components. 35, 108

PCA Principal component analysis. 11

PCR Polymerase chain reaction. 4, 34

PD Parkinson's Disease. v, 4, 19, 24, 33, 36, 37

PDD Parkinson's Disease Dementia. v, 19, 20, 24, 33, 36, 37, 40

PDE1A cyclic nucleotide phosphodiesterase 1. 45

PGC Psychiatric Genomic Consortium. 107

PINK1 PTEN-induced putative kinase 1. 28, 29, 60, 92

PMI post-mortem interval. 37, 43, 138

PNS peripheral nervous systems. 17

POLR3B RNA Polymerase III Subunit B. 99

PPP1R1B Protein Phosphatase 1 Regulatory Inhibitor subunit 1B. 56

PPP2R2B Protein Phosphatase 2 Regulatory Subunit Beta. 56

PRKN parkin RBR E3 ubiquitin protein ligase. 28, 29

PRS Polygenic risk score. 48

PSEN1 Presenilin 1. 29, 92

PSEN2 Presenilin 2. 29, 92

QC quality control. 41, 74, 104

RBD rapid eye movement sleep behaviour disorder. 33

RIN RNA integrity number. 34

RNA Ribonucleic acid. 4

RNA-Seq RNA Sequencing. 4, 8, 12, 110

rno6 Rattus norvegicus reference genome. 104

rRNA ribosomal RNA. 140

SCDE Single-cell differential expression. 12

SCN1A sodium voltage-gated channel alpha subunit 1. 18, 56, 124

SCN2A Sodium Voltage-Gated Channel Alpha Subunit 2. 124

scRNA-Seq Single-cell RNA Sequencing. 1, 4, 6, 8, 10, 12, 16, 76, 107, 111, 139

SCZ schizophrenia. 106, 116, 126

SFARI Simons Foundation Autism Research Initiative. 118, 119

SLC44A1 Solute Carrier Family 44 Member 1. 29

sLDSC stratified LD score regression. 15, 38

SNCA α-synuclein. 27, 29, 46, 60, 92

SNP single-nucleotide polymorphism. 1, 14

SNPs Single nucleotide polymorphisms. 106

snRNA-Seq Single-nucleus RNA Sequencing. 4, 6, 34, 39, 59, 70, 72, 74, 80, 82, 139

SON SON DNA binding protein. 99

SYNJ1 synaptojanin 1. 28, 60

t-SNE t-distributed stochastic neighbor embedding. 11

tauPLA Tau-proximity Ligation Assay. v, 72, 131

TBI traumatic brain injury. 31

TCF20 Transcription Factor 20. 99

TCF4 Transcription Factor 4. 99

TDP43 TAR DNA-binding protein 43. 93

TIMP2 Tissue inhibitor of metal- loproteinases 2. 45

TMEM175 Transmembrane Protein 175. 29

UCHL1 ubiquitin c-terminal hydroxylase 1. 28, 60

UMAP Uniform Manifold Approximation and Projection. 11, 35, 76, 82, 108

UMI unique molecular identifier. 35

USA United States of America. 18

USP9Y Ubiquitin Specific Peptidase 9 Y-Linked. 105

UTY Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked. 105

VPA Sodium valproate. 100–103, 106, 110, 111, 113–116, 118, 123, 125

VPS13A Vacuolar Protein Sorting 13 Homolog A. 60

VPS13C vacuolar pro- tein sorting 13C. 28

VPS35 vacuolar protein sorting 35. 28

WebGestaltR WEB-based GEne SeT AnaLysis Toolkit. 77, 106

WHO World Health Organization. 19

WHR Waist-to-hips ratio. 116

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the past few decades, interdisciplinary approaches to interrogate complex diseases have transformed the field of neuroscience research (Figure 1.1). Integrative genomics approaches using high throughput sequencing genomics data and computational tools are routinely used to identify disease-relevant biological and molecular pathways and characterise and study the function of vital organs [56, 77, 273]. Furthermore, associating disease-perturbed pathways with disease outcomes facilitates disease prevention, diagnosis and treatment [120, 204, 266].

For example, Jagadeesh et al.(2022) [133] used Single-cell RNA Sequencing (scRNA-Seq), epigenomic single-nucleotide polymorphism (SNP)-to-gene maps and Genome-Wide Association Studies (GWAS) summary statistics to infer cell type-phenotype relationships for neurological disorders, including γ-aminobutyric acid-ergic neurons in major depressive disorder. Furthermore, the authors also identified a disease-specific complement cascade process in multiple sclerosis. Similarly, Mathys et al.(2019) [205] used single-nucleus transcriptomes from the prefrontal cortex of late and early-stage AD samples to elucidate transcriptional changes in various cell-types. Additionally, the authors applied correlation analysis using gene expression and AD-related neuropathological traits using self-organising maps (unsupervised machine learning algorithms) to identify gene-trait correlation modules implicating genetic risk factors for both general cognitive function and AD.

In another example, Johnson et al.(2016) [142] applied system-level approaches to gene ex-

pression data from human hippocampus samples to draw connections between genetic factors that contribute to variation in cognitive ability (specifically, intelligence quotient intelligence quotient (IQ)) and those contributing risk to neurodevelopmental disease.

In this thesis, I will summarise three studies where I used integrative genomics approaches to elucidate some of the mechanisms underlying cognitive disorders arising from heterogeneous causes, including drug-induced neurodevelopmental impairments and neurodegeneration. In the current chapter, I will introduce the computational techniques used to study the transcriptome and briefly introduce the disorders of cognition under investigation. In Chapter 2 and Chapter 3, I will focus on diverse mechanisms that can cause dementia, including Lewy Body dementia and tau pathology, while Chapter 4 will focus on drug-induced neurodevelopmental impairments. In Chapter 5.1, I will summarise the convergent and divergent mechanisms underlying various types of dementia; lastly, I will discuss limitations and future work.

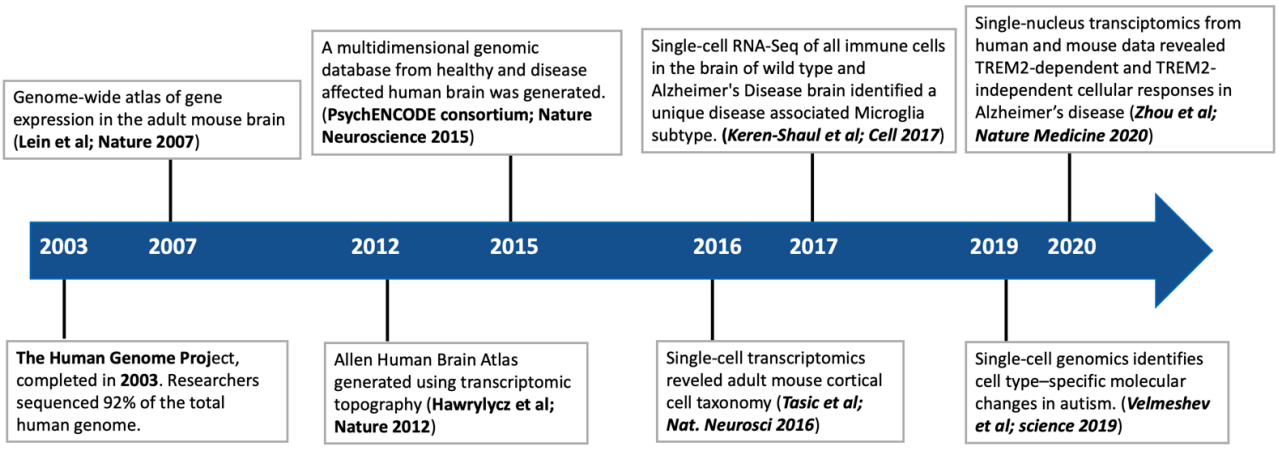


Figure 1.1: **Decades of Progress**. A timeline of some of the significant milestones in the field of genomics of neuroscience in the past few decades, highlighting the key breakthroughs and advancements that have shaped our understanding of the genetic basis of neurological disorders and brain function.

1.1 Genomics

Genomics is a branch of molecular biology that focuses on analysing and interpreting the entire set of genes, known as the genome, or a specific set of genes within an organism. Since the conception of the term "genomics" by Dr Thomas H. Roderick in 1986 [348], "omics" based biological research has been used to characterise and study biological processes that are vital for an organism to live. The omics era began in October 1990 after The Human Genome Project Human Genome Project (HGP) started. The project was launched by an international group of scientists with the aim of deciphering a reference human genome sequence to facilitate more effective approaches to studying human genetics. The result was generating an accurate reference sequence for each human chromosome (excluding large heterochromatic regions and a small number of gaps). This led to the HGP being instrumental in developing sequencing technologies for mapping and sequencing Deoxyribonucleic acid (DNA) [122]. Since the early 2000s, using *in silico* (computational) techniques for analysing large-scale genomics datasets has broadened our understanding of biology, including the human brain. Following the HGP project, several studies (Figure 1.1) have generated multiple biological data (including genomics, transcriptomics, and proteomics) from both human and model organisms. In 2007, the first genome-wide atlas of gene expression in the adult mouse brain was generated by Lein et al. [177]. While in 2015, a multidimensional genomic database of the human brain (from both healthy and diseased samples) was generated by the PsychENCODE consortium [5]. Although these databases have proven effective in medical research, many questions remain regarding brain disorders.

1.1.1 Bulk and Single-cell RNA-Sequencing

The first-generation DNA sequencing technology (FGS) was a chain termination method developed in 1977 by Sanger and Coulson [117]. The method uses gel electrophoresis to create sequencing ladders and performs base calling using fluorescent-based labelling techniques. Over the last decades, high-throughput next-generation sequencing (NGS) methods have evolved

to enable Ribonucleic acid (RNA) analysis by first converting RNA to complementary DNA (cDNA) by reverse transcription, followed by ligation of sequencing adaptors to the ends of the cDNA fragments and finally amplification of the fragments by Polymerase chain reaction (PCR) and generating RNA Sequencing (RNA-Seq) libraries [117]. Once libraries are prepared, DNA sequences are generated through a series of steps, which varies depending on the technology used [171, 277, 278, 368]. In general, libraries are first loaded onto a flow cell, where DNA fragments are immobilised and amplified, forming clusters containing multiple copies of the same sequences in close proximity. Next, a sequencing-by-synthesis process (using fluorescently labelled reversible terminators) occurs, involving iterative cycles of nucleotide incorporation and imaging for each cluster. The emitted fluorescence signal is detected to determine the sequence of each fragment. Following this, the raw fluorescence signal data undergo a data processing step, converting them into base calls that represent the DNA sequencing of each fragment. These base calls, accompanied by quality scores indicating the confidence in each call, are stored as sequence reads. The reads are then used in the down-stream analyses using several computational tools [171, 277, 278, 368].

Depending on the down-stream analyses (Figure 1.2), there are two widely used sequencing types: single-end short sequencing and paired-end longer sequencing, the latter of which is used to gain more comprehensive information about the transcriptome and aimed at analysis focusing on novel transcript identification, gene fusions, and point mutations as well as differential expression [181]. Bulk RNA-Seq data provides measurements of gene expression values that represent an average across all the different cell-types in the tissue being analysed. Until recently, most gene expression analyses were based on gene expression data from bulk tissue samples. However, developing methods to assay the transcriptomes of individual cells, using a method referred to as scRNA-Seq, offered the opportunity to resolve gene expression within individual cell-types. For example, early scRNA-Seq work identified novel transcriptionally distinct classes of microglia [93] and dopaminergic neurons [123] and showed that these cell-types were enriched in genetic risk to AD and PD, respectively. Initially, scRNA-Seq of the brain was limited to fresh tissue samples, however, the introduction of Single-nucleus RNA Sequencing (snRNA-Seq) enabled the application of scRNA-Seq advantages to frozen brain tissue samples

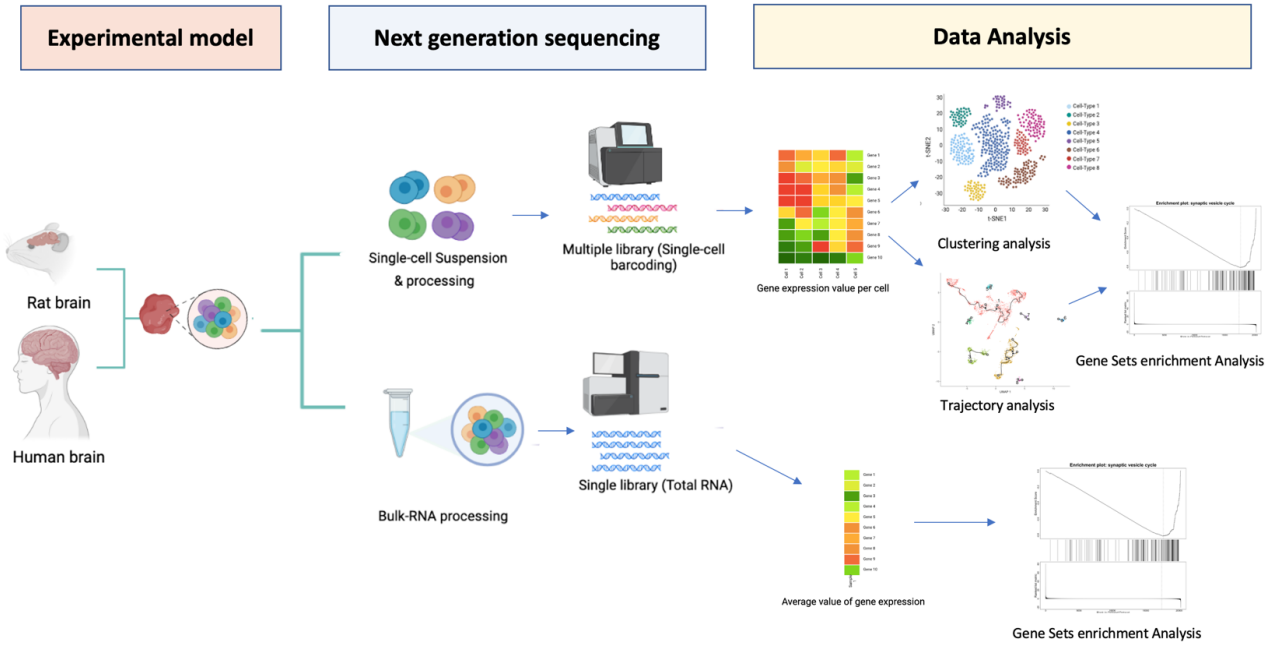


Figure 1.2: **Overview of Single-cell/Single-nucleus and bulk RNA Sequencing: Data Generation**The experimental design involves the selection of biological samples, library preparation, and sequencing on the appropriate platform. After sequencing, the raw data undergoes several pre-processing steps, such as quality control and alignment to a reference genome. For single-cell/single-nucleus RNA sequencing, data are further processed to identify and quantify gene expression at the single-cell level, including normalisation, gene filtering, and clustering. Finally, downstream analysis can include differential gene expression, pathway analysis, and cell-type identification. Abbreviation: RNA-Ribonucleic acid.

[166],[110]. Several studies have explored the comparability between the nuclear transcriptome and the transcriptome of the whole cell. These studies have consistently demonstrated the reliability of snRNA-Seq, providing strong evidence that snRNA-Seq accurately represents the RNA content at the tissue level [110],[165]. Furthermore, Habib et al. (2017) showed the expression profile of single nuclei to highly correlate with the average profile of single whole cells (Pearson r=0.87)[110]. Additionally, 98% of transcripts are represented in whole nuclei versus whole cells [319].

The challenges in detecting differences in gene expression levels associated with disease using bulk-tissue RNA-Seq are confounded by differences in cell-type abundances that may result from the disease itself or occur by chance due to sampling or other systematic differences between case and control samples. In contrast, at least in principle, scRNA-Seq and snRNA-Seq offers an opportunity to identify cell-type specific differences in gene expression associated with disease states and discover rare or uncharacterised cell-types and cell states in diseased compared to healthy tissues. Since then, both techniques have been widely used due to the rapid technological advancement that allowed parallel analysis of thousands of cells, which was also accompanied by the development of data analysis pipelines capable of analysing large amounts of biological data [181].

1.1.2 Analytical methods to study the transcriptome

Both bulk and scRNA-Seq technologies provide powerful platforms for comprehensive analysis of the whole transcriptome. Quantitative gene expression analysis is crucial for generating hypotheses about the molecular mechanisms underlying disease and physiological states, and several computational pipelines have been developed to carry out this task [162, 352]. The pipelines for data analysis (after mapping reads to a specific reference genome and gene expression quantification), in general, include three main sections (Figure 1.3) [162, 352, 322, 58, 316]:

(A) **Data pre-processing**, including (i) quality control of count data; (ii) integration of count data.
(B) **Data analyses**, including (i) dimensionality reduction; (ii) clustering.

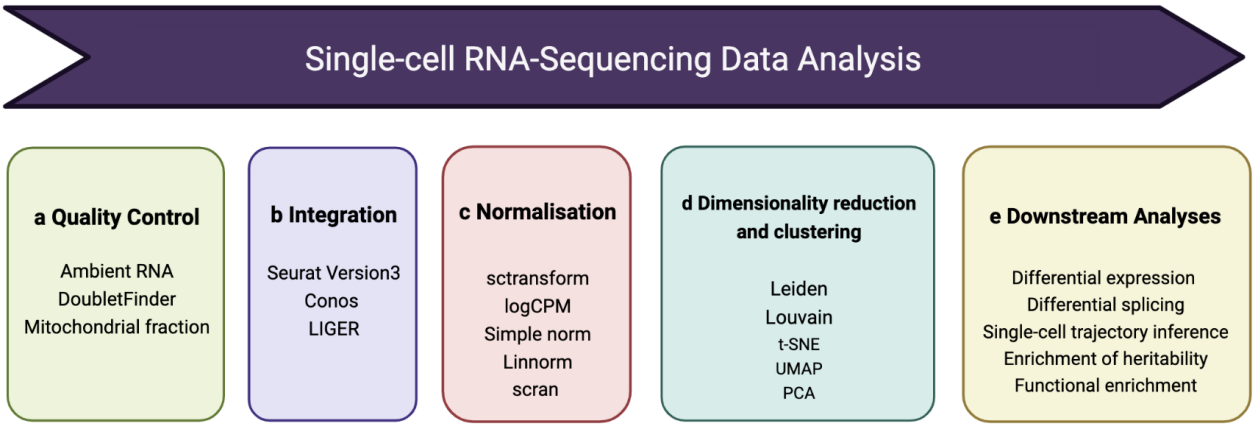


Figure 1.3: **An overview of single-cell/single-nucleus RNA sequencing data analysis.** The Single-cell/Single-nucleus RNA-Seq analysis pipeline refers to a series of computationally based analyses of transcriptomic data. The process involves several steps, including measuring the level of gene expression, checking the quality of the read counts, integrating the data, normalising the data, selecting specific features (most variable genes), reducing the dimensionality of the data, clustering cells with similar profiles, annotating the cells based on their characteristics (cell-type identification), and performing further analyses. The final set of analyses is particularly important as it helps researchers to distinguish between different cell populations, identify genes that are differentially expressed between different cell-types or conditions, and conduct various other analyses based on the transcriptomic profiles of the cells. Abbreviations: LIGER (Linked Inference of Genomic Experimental Relationships); CPM (counts per million);t-SNE (t-distributed stochastic neighbour embedding); UMAP (Uniform Manifold Approximation and Projection); PCA (Principal component analysis); RNA (Ribonucleic acid).

(C) **Downstream analyses**, including (i) differential splicing analysis; (ii) differential expression analysis; (iii) further downstream analyses to determine the functional consequences of differentially expressed or spliced genes. These analyses include (a) functional enrichment analyses, which test whether a set of genes is enriched for genes that contribute to certain biological processes, molecular function or cellular compartments beyond random expectation and (b) heritability enrichment analysis which tests whether a gene set (for example one consisting of genes differentially expressed between case and controls) is enriched for genes which contribute genetic risk for disease (heritability enrichment analysis) and (c) single-cell trajectory inference to explore the underlying path gene expression follows in a dynamic biological process such as between disease or developmental states.

1.1.3 Single-cell data pre-processing and analysis

Data pre-processing

A crucial requirement for a successful RNA-Seq study is a well-defined and meaningful experimental design. The design should start with a clear understanding of the biological questions to be addressed, which will influence the choice of library type, the number of biological replicates per condition (which influences statistical power), sequencing depth and steps required to minimise biases and/or confounding batch effects.

The preliminary steps for data pre-processing of RNA-Seq datasets include a quality check of raw reads outputted from the sequencer. For bulk RNA-Sequencing datasets, this is done using tools such as FastQC [10] and STAR [75] to map reads to a reference genome, adapter trimming, and generate counts. For scRNA-Seq, data pre-processing analysis (Figure1.3a-b) steps can be done using a tailored pipeline such as CellRanger [363](which was developed to carry out quality checks of reads, mapping and count generation of single-cell datasets) or other computational tools tailored for each step separately [209, 270]. As part of the quality control initial steps for scRNA-Seq datasets, removing potentially 'empty' cells from the data matrices is a crucial step. These empty cells, also called ambient RNA-containing cells (cell-free transcripts in the solution in which the cells are suspended [191, 196, 369]), are among the main issues of using droplet-based scRNA-Seq assays. Their presence in the dataset could lead to systemic biases in downstream analysis [191]. In order to resolve this, Fleming et al.(2019) [369] used a deep generative model to learn the background RNA profile and separate cell-containing droplets from empty ones.

In some cases, single-cell datasets are generated using different technology platforms with differences in equipment, capturing time, reagents, and even handling personnel. These differences could introduce biases due to incomplete library sequencing, capture efficiency, differences in technology platforms or amplification bias (Figure1.3b), making samples incomparable [279].; therefore, batch-specific systemic effects need to be removed prior to downstream analysis [113]. To address this, several workflows have been developed. For example, to handle scRNA-Seq data, a mutual nearest neighbours (MNNs) based algorithm is used to align the datasets into

shared space. The algorithm first identifies similar cells (mutual neighbours) between batches. The difference (batch effect) between the cells is then quantified and used as a scale to merge the datasets (batches). The degree of difference is also used to assess the batch effect's strength. Recently, a computationally less demanding approach known as fastMNN has been developed to compute the neighbours in dimensionally reduced space [113]. Among recently developed integration tools are:

(i) **Clustering On Network Of Samples (Conos)**, a method which uses a graph-based approach to construct a global graph (connecting all cells across several datasets). First, the algorithm identifies MNNs between cells in different datasets. Next, MNNs align the datasets and integrate them into a unified representation. Further, Conos allows the identification of clusters found across all datasets and those uniquely found in each dataset [19].

(ii) **Seurat Version 3** provides a method which aligns data batches using canonical correlation analysis (CCA) based approaches. Briefly, shared correlation structures are identified across datasets using CCA (which projects the data into a subspace). Next, MNNs are computed in the subspace and used as anchors to correct the data [299].

(ii)**Linked Inference of Genomic Experimental Relationships (LIGER)**, an integration tool which uses a graph-based approach to integrate datasets. Briefly, the algorithm first constructs a graph using all cells, where each node represents a cell and the edges represent the similarity between cells based on their expression profiles. Next, the algorithm uses a graph-matching method to integrate the datasets by finding a common space in which cells from different datasets are aligned based on their expression profiles. After integration, joint clustering is performed to identify cell types conserved across the different datasets [338].

Data analysis

Furthermore, to ensure comparability of expression measurements across different cells in downstream analysis, it is crucial to address biases introduced due to technical variations, such as differences in sequencing depth, gene length, GC content, and dropout events. These biases can obscure underlying biological differences between samples, making it challenging to accurately

interpret biological signals. To overcome this limitation, adjusting the expression measurements to make them proportional to the true expression levels (to normalise the data) before using raw counts in the downstream analysis is necessary [192, 41, 8].

Generally, the normalisation process includes either a linear regression approach (to estimate size factors) or a non-linear approach using parametric modelling on count data [8, 41, 112]. In both approaches, absolute counts are turned into ratios (i.e., how many transcripts of gene A are there per 10,000 total counts), and then the relative number (normalised value) is transformed to reduce the skewness or regress technical variation. Commonly used normalisation methods in the scRNA-Seq analysis include:

i) **LogNorm**, a method that is based on log transformation in which each gene is first transformed by taking the logarithm of the count values, followed by standardization of the value across all cells by subtracting the mean expression value and dividing by the standard deviation [41].

(ii) **counts per million (CPM)**, a widely used normalisation method which accounts for differences in sequencing depth. CPM is computed by dividing the raw count for each gene by the total number of reads in the sample, followed by multiplying (scaling) the value by a million [8].

(iii) **sctransform**, a method specifically designed for UMI-based scRNA-Seq data with high levels of technical noise or batch effects. Unlike other normalisation methods, which make assumptions about the data distribution, sctransform uses the Bayesian framework to model the distribution of technical noise in each cell. Briefly, sctransform fits a generalised linear model for sequencing depth as the explanatory variable, and UMI counts as the response variable. Next, each count is transformed into a Pearson residual using the parameters computed by the model. The transformed count is then used as a normalised value [112].

Moreover, dimensionality reduction techniques are applied in scRNA-Seq data analysis to mitigate noise that could impede the identification of meaningful patterns or structures [302]. By transforming the high-dimensional scRNA-Seq data into a lower-dimensional space, dimensionality reduction techniques retain the essential features that contribute to variation in the

datasets while removing irrelevant features and noise, thus, improving the accuracy and interpretability of biological signals.

Principal component analysis (PCA) is a dimensionality reduction technique widely used to visualise biological or technical variation within a dataset [143]. The PCA method takes large input data (gene counts) and reduces the dimensions to a set of linearly transformed dimensions while maintaining as much of the data's variation as possible. The transformed dimensions are plotted as a two-dimensional plot. The plot can also be used in identifying technical and biological outliers. However, for complex structured datasets such as (single-cell datasets), non-linear combination methods are used, such as t-distributed stochastic neighbor embedding (t-SNE) [183] and Uniform Manifold Approximation and Projection (UMAP) [20]. While both methods are used to visualise complex datasets, UMAP estimates low dimensions of the data using cell-cell nearest neighbour network to characterise the global structure of the dataset, while t-SNE reduces high dimensional dataset to low dimensional graph but focuses more on capturing local structure instead of the global structure.

In general, dimensionality reduction is applied prior to clustering, as reducing the dimensionality of high-dimensional data can help improve the performance and interpretability of clustering algorithms. The main goal of clustering is to group cells based on the similarity of their gene expression profiles, which can reveal cell-types, states, or sub-populations that may be present in the dataset. The two commonly used algorithms, Louvain and Leiden are based on K-Nearest Neighbour (KNN) graph [26, 313]. These (graph-based) clustering methods compute Euclidean distance on the PC reduced expression space to connect each cell to its nearest (k most similar) cell in the same dataset, and each cell is represented as a node in a graph. Once clusters are defined, known cell-type specific marker genes are used to identify the cell-type and downstream analyses are carried out on a cell-type basis.

The main difference between the Louvain and Leiden systems is their approaches to community detection. While Louvain starts with individual nodes in different communities and repeatedly combines them to optimise modularity scores, Leiden includes a refinement phase that evaluates the quality of each individual node migrations across communities. Compared to Louvain, this refining phase enables Leiden to create more accurate partitions, particularly in networks with

highly heterogeneous community sizes or noisy data.

Data Interpretation

After clustering and identifying cell-types, several downstream analyses are carried out, including differential expression analysis, gene sets enrichment analysis and enrichment of heritability.

(a) Differential expression analysis

Differential expression analysis methods seek to quantify differences in gene expression between conditions and can be applied to single transcripts or genome-wide gene expression datasets derived from microarray [288], bulk RNA-Seq or scRNA-Seq datasets [8, 88]. Count data are usually normalised to enable direct comparison of gene expression between samples/groups, which is then analysed using a statistical method. Current methods used for bulk RNA-Seq differential analysis can be grouped into three main categories: (i) non-parametric methods (ii) Poisson or negative binomial model-based and (iii) linear models [290]. Parametric methods assume count distributions follow a specific distribution (log-normal, negative binomial, Poisson, or empirical bayes). For example, edgeR [257], a method which uses a parametric model (negative binomial) to model the counts. In contrast, non-parametric methods do not rely on such assumptions; therefore, they tend to be more robust when the expression of a gene deviates from a distribution assumed by the model [290].

Unlike bulk RNA-Seq-based methods, single-cell-based methods can theoretically characterise complex responses to biological or experimental perturbations. However, due to the heterogeneity, sparsity, small library sizes, drop-out events (due to the low capture efficiency of RNA molecules, some transcripts can be highly expressed in one cell and completely missing in another cell of the same population), and high noise level found in single-cell data, it is challenging to produce biologically accurate results using bulk-RNA Seq based methods [127, 190, 88]. For this reason, new methods have been proposed to address the challenges of drop-out events and multimodal expression values. Among the most commonly used methods, Single-cell differential expression (SCDE) [153] and Model-based Analysis of Single-cell Transcriptomics (MAST) [89] use a two-part joint model to identify differentially expressed genes.

MAST is a tool tailored for scRNA-Seq data analysis. The method was developed to address the challenges associated with scRNA-Seq data including, bimodality in expression levels; technical assay variability, for example, messenger RNA (mRNA) quality; and extrinsic biological factors including, differences in cell sizes. MAST is based on a two-part generalised linear model that models the discrete expression rate of each gene across cells (using logistic regression) and the continuous positive gene expression mean (using the Gaussian linear model) in parallel. This allows a joint estimate of background and treatment effects. Further, MAST uses the fraction of genes detectably expressed in each cell termed Cellular Detection Rate (CDR) as a covariate to adjust for technical factors, including the dropout and biological factors, including cell volume.

Wang et al. 2019 [333] recently published a comparative study of Differential Expression Analysis (DEA) methods used, in which they observed methods with high true positive rates with low precision, while methods with high precision showed low true positive rates. Therefore, the authors concluded that precautions must be taken when choosing a method, and it is important to have a trade-off between true positive rates and the precision of identifying differentially expressed genes.

(b) Functional enrichment analysis

Functional enrichment analysis, including Gene-set enrichment analysis (GSEA) and overrepresentation analysis, is carried out to examine whether identified DEG sets are meaningful to biological or molecular mechanisms and to generate hypotheses regarding the underlying biological consequences of experiments. GSEA [301] has been used in downstream analysis since the year 2001 when Gene Ontology (GO) [61] was first launched to facilitate the interpretation of DEGs generated by cDNA microarrays. GO consists of a hierarchical structure of terms related to biological processes, cellular components or molecular functions and their relationships to lists of annotated genes. Gene annotations were manually assigned from experimental analysis or automated/computational methods. GO enrichment analysis is carried out to test if GO terms are under or over-represented within the DEGs sets or gene set of interest [301, 182].

(c) Enrichment of heritability

Heritability enrichment analyses are statistical methods used to test whether a set of genes (or genomic regions) are enriched for heritability to a specific trait or physiological phenotype beyond random expectation. Since gene expression may vary both as a cause and a consequence of disease, statistically significant enrichment of heritability is usually taken to suggest that the gene set has a proximal (i.e., causal) relationship to the phenotype under investigation. These forms of analysis can be done using common genetic variation SNP data from GWAS[324] or rare variant analysis [176].

There are various methods available for analyzing rare variants; one such approach is the Combined Multivariate and Collapsing Rare Variant Analysis (CMC) [179]. The method tests for association between rare variants and a trait by first collapsing variants within a genomic region, then comparing the frequency of the collapse variants between cases and controls. Further, it can be applied to genomics regions, including exons, introns and regulatory regions. Further, the heritability enrichment score (HES) for each region, as defined by the method, measures the extent to which a genomic region is enriched for heritability relative to the genome-wide average [179].

In a recent study, the authors performed a rare variant heritability enrichment analysis to identify genomic regions enriched for heritability in schizophrenia. By meta-analysing the whole exomes of over 121,000 cases and controls, the authors implicated several rare coding variants in 42 genes as conferring risk for schizophrenia (odds ratios of 3–50, $P < 2.14 \times 10^{-6}$ for 10 genes; and 32 genes at a false discovery rate of < 0.05) [282].

Common variants are typically used in heritability enrichment analyses because they are more prevalent in the population than rare variants, making them easier to study. There are several methods for conducting heritability enrichment analysis using common variants. Multi-marker Analysis of GenoMic Annotation (MAGMA) [43], and Linkage Disequilibrium Score Regression (LDSC) [40] are widely used methods to estimate genetic correlation. MAGMA uses a gene-level regression approach by first generating gene-level p-value as the mean of association of variants in the gene (using GWAS summary statistics and Linkage Disequilibrium (LD) corrected) and

then calculates continuous p-values based on the association of the mean (overall) p-value of the gene with that of the gene of interest. In other words, the gene-set analysis tests if the genes in a specific set are more strongly associated with a trait than other genes.

Similarly, LDSC uses regression analysis to quantify the association between LD and test statistics of variants (using GWAS summary statistics). The final LD score of a variant is the sum of LD r^2 measured with all other variants. Unlike MAGMA, LDSC produces variants-level heritability estimates to partition heritability into separate categories. Additionally, an extension of LDSC known as stratified LD score regression (sLDSC), partitions heritability using functional annotations by considering the genetic linkage between variants.

Further, LDSC can distinguish between confounding factors such as population stratification and polygenicity. Both methods use GWAS genotype data to extract summary statistics about the causal variant. This is because using well-powered GWAS data rather than other sources of information (for example, using information from studies with well- characterised pathways and well-studied genes) prevents biases towards specific genes. Typically in RNA-Seq studies, enrichment of heritability is calculated using DEG sets as a set of interest to see if these genes can be genetically linked to a disease or phenotype for which the GWAS summary statistics were generated.

(d) Single-cell trajectory Analysis

One of the challenges in the analysis of the scRNA-Seq dataset is that the data only provides a snapshot of gene expression at a particular point in time. Throughout life or during development, cells transition from one state to another, each cell in different states expressing different sets of genes or genes at different levels of expression. This may be due to the cells being at distinct points of a dynamic process, for example, cell cycle or cell differentiation, or some cells in a tissue sample reflecting a particular developmental or disease state [263].

Trajectory inference methods, also known as pseudotime analysis, aim to order cells (using scRNA-Seq data) along a trajectory or continuum from one state to another based on similarities in the expression patterns of individual cells [45]. The trajectories (branches) are conceptualised as cellular "decisions" and can be linear, cyclic or tree-branch. This is because there are several lineages within a population of cells, each lineage steaming from a common initial group. Therefore, each branch (trajectory) can be used to investigate several cellular processes, including cell differentiation and development as well as transitions between disease states [45, 144].

A recently developed method, Monocle [316], uses scRNA-Seq data to construct single-cell trajectories. The algorithm first uses an unsupervised algorithm to learn the sequence of gene expression changes a cell goes through as part of the biological process of interest. Then, each cell is placed at its specific position in the trajectory. Monocle also uses regression analysis or graph autocorrelation analysis to identify genes that change expression as a function of the inferred pseudotime-a quantitative measurement of progress. Furthermore, monocle identifies co-regulated modules of differentially expressed genes [251]. In Monocle, modules refer to sets of genes that exhibit co-expression or co-regulation during a specific biological process. These modules are identified using unsupervised clustering techniques that group genes with similar expression patterns. By grouping co-expressed or co-regulated genes, Monocle enables the identification of key regulatory pathways and molecular mechanisms underlying cellular processes [46].

1.2 Disorders of cognition

The Oxford English Dictionary defines cognition as "the mental action or process of acquiring knowledge, thought, experience and the senses". Cognition encompasses all aspects of intellectual functioning (termed domains). The Diagnostic and Statistical Manual of Mental (DSM-5), 5th Edition classifies these domains into six key categories: executive function, language, complex attention, learning and memory, perceptual-motor control, and social cognition [116]. These main cognitive domains play a key role in diagnosing neurological disorders. Assessments are typically carried out by several tasks aimed at each ability area. The main aim of the task is to measure and differentiate between different domains; a pattern of impairment in specific domains or across domains is often used to underpin a clinical neurological or neuropsychiatric diagnosis [116]. Neurocognitive disorders are a group of heterogeneous disorders encompassing a broad spectrum of clinical phenotypes. They may arise through multiple physiological (e.g., ageing) and pathological processes (e.g., neurodevelopmental abnormalities, genetic mutation, drugs, brain injury, epilepsy, neurodegeneration) [208, 148]. Therefore, they are classified in multiple ways and levels, such as the nature of the cognitive impairment or the aetiology.

1.2.1 Neurodevelopmental influences on cognition

Neurodevelopment, i.e., the development of the nervous system, is a complex and precisely regulated process involving neural tube formation, neural and glial proliferation and neuronal migration and organisation. A deficit introduced at any critical stage of development could lead to neurodevelopmental disorders [238]. Neurodevelopmental Disorders (NDDs) are neurological disorders arising from a defect in neurodevelopment and may affect both the Central nervous systems (CNS) and peripheral nervous systems (PNS), respectively [256]. Examples of CNS NDDs include learning disabilities, Autism Spectrum Disorder (ASD), intellectual disability (ID), Attention-deficit/hyperactivity disorder (ADHD), cerebral palsy, conduct disorders, vision and hearing impairments and epilepsy. Individuals with these types of disorders may experience difficulties across a range of neurological functions such as memory and learning, motor skills, language and speech[238]. Treatments often involve a combination of pharmaceuticals as well

as occupational and physical therapy. The prevalence of NDDs (as defined by DSM-5)58 is between 0.7-17% under 18 years old worldwide (spanning Africa, Asia, Australia, Europe, Latin America and the United States of America (USA)) [92]. The reported prevalence rates of individual NDDs include ADHD (5-11%), ASD (0.7-3%), and epilepsy (0.5%) [92]. Most NDDs have multiple risk factors, including genetic, biological, environmental (e.g., alcohol, illicit drugs, environmental contaminants), psychological and social. However, recent studies have shown some NDDs (e.g., IDs) are associated with mutations in specific genes. For example, DNA variants in sodium voltage-gated channel alpha subunit 1 (SCN1A), a gene that provides instructions for a protein responsible for the flow of sodium into neurons, are known to cause Dravet Syndrome, a severe form of epilepsy and intellectual disability [83].

Intellectual Disability

Intellectual disability (ID) is a neurodevelopmental disorder affecting around 1–3% of the world's population [18]. DSM-5 defines intellectual disability as a neurological condition characterised by (i) cognitive impairment in intellectual functioning (domains) such as problem-solving, learning (from experience as well as academic learning), vocabulary, reasoning, planning, abstract thinking, and judgement (ii) deficits in adaptive functioning including self-care, social and interpersonal skills (iii) early onset of these deficits (during the developmental period) [318]. Previously, ID classification was based on Intelligence quotient (IQ) scores. For example, an IQ between 50-70 is mild mental retardation, and IQ < 50 is used to define severe mental retardation [318, 129]. The severity of ID is assessed based on the level of support an individual has (based on the severity of the adaptive functioning deficits in the social, conceptual and practical domains). As with many other NDDs, the main causes of intellectual disability are not well defined. Among many contributing factors, including de-novo rare deleterious mutation [197, 325] as well as exposures to environmental contaminants (such as lead and mercury), which have been associated with intellectual disorders [207].

1.2.2 Neurodegenerative influences on cognition

Neurodegenerative Diseases (NDs) are ultimately debilitating conditions associated with progressive neuronal loss. For CNS NDs, symptoms can range from a progressive motor dysfunction, as in amyotrophic lateral sclerosis (also known as motor neurone disease) or PD, to cognitive deficits in memory and executive function, as in various forms of dementia, including AD, DLB and PDD (the latter two also referred to as "Lewy Body Dementias"). Research has shown that age and a combination of genetic and environmental factors contribute to the risk of developing a neurodegenerative disease [124, 343, 36, 168].

Dementia is a term used to describe a deterioration in cognitive function beyond what is considered to be the biological consequence of ageing [235]. Dementia has many causes, unified around damage and loss of neurons. Dementia can affect several brain functions, including learning and memory, executive function, comprehension, calculation, movement and language. According to a recent report published by World Health Organization (WHO), approximately 55 million people live with dementia worldwide, and 10 million new cases are added every year. The social and economic impact of dementia is expected to increase with an increasing population. Although dementia is considered an age-related disease, it does not exclusively affect older people, and about 9% of cases are early onset-meaning the onset of symptoms began before the age of 65 [235], AD is the most common cause of dementia and accounts for 60-80% of cases. This type of dementia affects behaviour, thinking and memory. AD changes usually in the part of the brain responsible for learning, the entorhinal-hippocampal system [267]; therefore, early AD symptoms include difficulty remembering newly learned information. As the disease progresses, it affects other areas of the cerebral cortex involved in navigation, spatial orientation, behaviour, and personality [234].

PD is the second most common neurodegenerative disease. Typically, the basal ganglia and the substantia nigra are among the first regions to be involved. These brain regions play a key role in movement, leading to early symptoms that include rigid muscles, slowed movement (bradykinesia), lack of facial expression and difficulty with balance and coordination [218]. As brain

changes caused by PD gradually spread, other brain areas begin to be affected. For example, the individual may experience cognitive decline, including fluctuation in attention, memory and executive function [218]. By convention, if the cognitive decline develops more than one year after a PD diagnosis, then the individual is diagnosed with PDD [81]. However, if the cognitive decline is a presenting symptom or develops within one year of the onset of the motor symptoms characteristic of Parkinsonism, then the individual is diagnosed with DLB [210]. Therefore, the timing between the identification of dementia and the initiation of parkinsonism symptoms is used to determine if the dementia phenotype is diagnosed as PDD or DLB. The pathological hallmark of PD, DLB and PDD is the abnormal deposits of alpha-synuclein known as "Lewy bodies", leading to these three disorders as Lewy Body Diseases (LBDs). The mechanisms underlying LBDs remain largely unknown but include genetic [245]and environmental risk factors [245].

1.3 Overview of Thesis

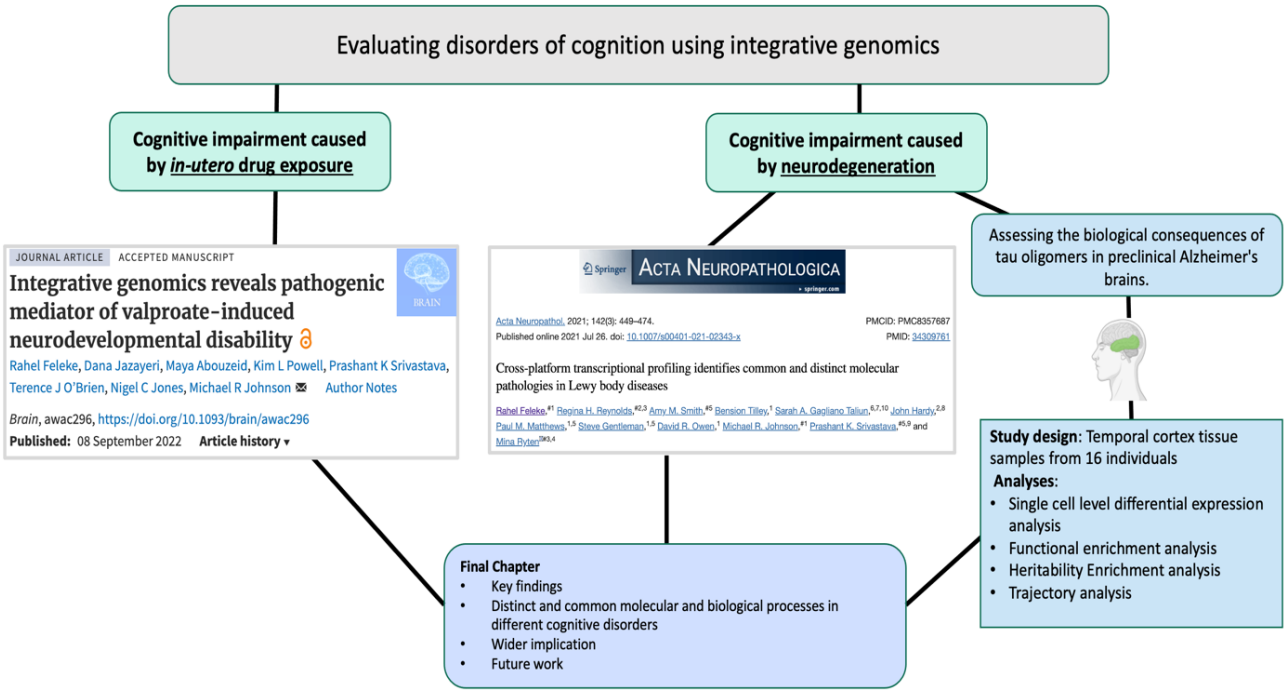


Figure 1.4: **An overview of the chapters of the thesis.**The chapters are divided into three main sections. Section A (2) focuses on cognitive impairments caused by *in utero* drug exposure. Section B (3 and 4) focus on cognitive impairments caused by neurodegeneration. Section C (5.1) is the discussion and conclusion chapter.

1.4 Publications

1. Alexander Haglund, Verena Zuber, Yifei Yang, Maya Abouzeid, **Rahel Feleke**, Jeong Hun Ko, Alexi Nott, Ann C. Babtie, James D. Mills, Louwai Muhammed, Liisi Laaniste, Djordje O. Gveric, Daniel Clode, Susanna Pagni, Ravishankara Bellampalli, Alyma Somani, Karina McDade, Jasper J. Anink, Lucia Mesarosova, Eleonora Aronica, Maria Thom, Sanjay M. Sisodiya, Prashant K. Srivastava, Dheeraj Malhotra, Julien Bryois, Leonardo Bottolo and Michael R. Johnson. "Single-cell Mendelian randomisation identifies cell-type specific genetic effects on human brain disease and behaviour." bioRxiv (2022): 2022-11.

2. **Rahel Feleke**, Dana Jazayeri, Maya Abouzeid, Kim L. Powell, Prashant K. Srivastava, Terence J. O'Brien, Nigel C. Jones, and Michael R. Johnson. "Integrative genomics reveals pathogenic mediator of valproate-induced neurodevelopmental disability." Brain 145, no. 11 (2022): 3832-3842.

3. **Rahel Feleke**, Regina H Reynolds, Amy M Smith, Bension Tilley, Sarah A Gagliano Taliun, John Hardy, Paul M Matthews, Steve Gentleman, David R Owen, Michael R Johnson, Prashant K Srivastava and Mina Ryten. "Cross-platform transcriptional profiling identifies common and distinct molecular pathologies in Lewy body diseases." Acta Neuropathologica 142 (2021): 449-474.

4. Alina Schidlitzki, Pablo Bascuñana, Prashant K Srivastava, Lisa Welzel, Friederike Twele, Kathrin Töllner, Christopher Käufer, Birthe Gericke, **Rahel Feleke**, Martin Meier, Andras Polyak, Tobias L Ross, Ingo Gerhauser, Jens P Bankstahl, Michael R Johnson, Marion Bankstahl and Wolfgang Löscher. "Proof-of-concept that network pharmacology is effective to modify development of acquired temporal lobe epilepsy." Neurobiology of disease 134 (2020): 104664.

Chapter 2

Single-nucleus transcriptomics identifies common and distinct molecular pathologies in Lewy body diseases

Abstract

Parkinson's disease (PD), Parkinson's disease with dementia (PDD), and Dementia with Lewy bodies (DLB) are neurodegenerative disorders collectively known as Lewy body diseases (LBDs). This study aimed to comprehensively profile the transcriptomes of LBDs using single-nucleus RNA-Sequencing. Analysis of anterior cingulate cortex samples from neurologically healthy controls, PD, PDD, and DLB (n=7 per group) cases revealed significant transcriptional alterations across multiple cell types, providing insights into the commonalities and distinctions among these disorders. The selective vulnerability of medium spiny neurons (MSN) in DLB was observed, whereas PD and PDD subjects exhibited MSN-like neurons. Pathway analysis highlighted shared and distinct perturbations in biological processes, with (i) synaptic vesicle transport and oxidative stress response being uniquely down-regulated in the dementias (PDD and DLB); and (ii) up-regulated genes in PD were associated with metabolic processes, mRNA splicing, and autophagy. Furthermore, genetic associations indicated a link between up-regulated genes in astrocytes and oligodendrocyte precursor cells with PD age of onset risk, while microglia showed enrichment for Alzheimer's disease (AD) risk. However, down-regulated genes in LBDs did not exhibit significant genetic associations. The findings provide valuable insights into the transcriptomic landscape of LBDs, emphasizing the unique vulnerability of MSN in DLB and revealing shared and distinct molecular pathways. The study contributes to our understanding of LBDs and lays the groundwork for targeted therapeutic interventions. Nonetheless, limitations in sample composition and data analysis tools should be considered when interpreting the results.

2.1 Introduction

First described in 1817 by an English surgeon James Parkinson [101], Parkinson's disease (PD) is a progressive neurodegenerative disorder reported to affect approximately 10 million people worldwide [229]. Clinically, PD affects both the motor system [102], which causes symptoms collectively known as parkinsonism, including rigid muscles, slowed movement (bradykinesia) and lack of facial expression and non-motor systems across several cognitive domains, including executive and visuospatial functions. Neuropathologically, PD is characterised by the presence of Lewy bodies, eosinophilic intraneuronal inclusions first described in 1912 by Fritz Heinrich Lewy [82]. It is estimated between 50 to 80% [2] of people with PD eventually develop dementia [7], with an average time of about ten years from the onset of parkinsonism (motor symptoms) to the development of dementia. This form of PD is known as PD with Dementia (PDD) [7, 244].

Dementia, as described in previous Chapters 1.2.2, is a term used to describe a group of symptoms characterised by a progressive decline in attention, memory and behaviour, and frequently also involves autonomic dysfunction [242]. After Alzheimer's disease, the second most common form of dementia is Dementia with Lewy Bodies (DLB) [146]. As with PD, DLB is characterised by the presence of Lewy bodies, now understood to represent aggregates of alpha-synuclein protein [82, 146]. The clinical features of DLB and PDD are similar due to the overlap of cognitive impairments and diagnostic features. For example, the core diagnostic features of DLB, such as parkinsonism, fluctuating attention and cognition, are also considered core features of PDD [137]. This overlap led to a theory suggesting a continuous clinical spectrum for Lewy Body Dementias, which collectively encompasses both PDD and DLB (not to be confused with Lewy Body Diseases (LBDs), which includes PD) [62].

The distinction between DLB and PDD clinically is the time of onset of the parkinsonism relative to the cognitive impairment, such that if cognitive impairment begins within 1 year of the onset of parkinsonian motor symptoms, then clinically, the symptoms are described as DLB. However, if cognitive impairments happen after 1 year of the onset of parkinsonism, the

symptoms are clinically defined as PDD [137, 62, 12]. With the recognition of the severity of the disease and the frequent difficulties in distinguishing PDD from DLB, there have been growing efforts to identify distinct clinical, neuropathological and genetic features of LBDs [12].

2.1.1 Neuropathological features of Lewy Body Diseases

Since the discovery of the chemical composition of Lewy Bodies by Spillantini et al. (1997) [292], PD has been described as a disorder of α-synuclein (α-syn), a neuronal protein responsible for regulating synaptic plasticity, promoting the synaptic organisation and vesicle transportation [297]. The two major neuropathological features of PD are (i) accumulation of intraneuronal α-synuclein-containing aggregates in the form of Lewy bodies and Lewy neurites, termed Lewy bodies pathology and (ii) loss of dopaminergic neurons in the substantia nigra [159].

The Braak staging system assesses the stage of α-synuclein pathology in PD. First introduced in 2003, Braak and colleagues developed six stages of Lewy bodies pathology in PD based on the idea that the temporal spread of the pathology originates in the brainstem (stages 1-2) and then spreads through the midbrain, basal forebrain, subcortical, temporal mesocortical regions (stages 3-4), and eventually spreads to the cerebral cortices where it appears in the anterior cingulate mesocortex and spreads to other cortical regions (stages 5-6) [35, 34, 317]. This predictable spread of Lewy bodies pathology supports the hypothesis that pathology propagates through the brain via a prion-like process [328]. Although this holds in most post-mortem cases, in which there is a predictable sequence of lesions, results from other studies have shown that the severity of symptoms and the duration of parkinsonism does not always correlate with the Lewy bodies density and distribution [237].

For example, Jellinger (2009) reported that in some autopsy studies, up to 43% of PD cases do not follow the proposed pattern of progression [136]. In light of this, McKeith et al. (2017) [211] updated their diagnostic guidelines for assessing Lewy bodies pathology based on the predominance of pathology in specific regions, including the neocortical regions. Further, Levernez et al. (2008) [178] increased the specificity of this method by reducing the number of regions assessed and adding an amygdala-predominant category. PD, PDD and LBD can be viewed

as existing on a spectrum of diseases related to each other via Lewy body pathology [137, 62]. In autopsy-based studies involving PD and PDD cases, the prevalence and load of neocortical α-synuclein were higher in PDD cases than in PD cases [287]. Similarly, α-synuclein loads in cortical regions (including the entorhinal, parietal, frontal, cingulate, and temporal cortex) are observed to be the highest in DLB and the lowest in PD, and intermediate in PDD cases [119, 27]. These observations are consistent with the dementia arising as a result of spread of Lewy bodies pathology through the cerebral cortex.

In addition to being conceptualised as alpha-synucleinopathies, Lewy body diseases (PD, PDD and DLB) are frequently associated with, and may be modulated by, comorbid AD-related pathology (hyperphosphorylated tau and amyloid-β plaques) [256]. For example, Jellinger et al. (2018) [137] observed amyloid-β (Aβ) burden was significantly higher in cortical regions in DLB than in PDD. Similarly, Kalaizakis et al. (2013) [145] showed that Aβ lesions were increased in PDD compared to PD in the entorhinal cortex. Further, studies have revealed an intriguing finding that links amyloid-β and tau neurofibrillary pathology to a reduced timeframe between the initiation of parkinsonism symptoms and the development of dementia [100, 57]. Furthermore, a cortical load of Lewy pathology was a good predictor of dementia in PD cases. Therefore, it has been suggested that the neuropathologies of tau, Aβ, and α-syn may all synergistically contribute to cognitive decline in the dementias with Lewy bodies [55].

2.1.2 Genetics of Lewy Body Diseases

Currently, PD is identified as a genetically heterogeneous disorder with two subtypes: (i) monogenic (variation in a single gene) familial PD with mendelian inheritance, a type of biological inheritance in which a single gene of high penetrance confers risk for a trait with dominant and recessive alleles inherited from either parent; (ii) a sporadic form arising from complex inheritance with less familial aggregation [157, 25]. In contrast, sporadic forms of PD are thought to result from a contribution of several factors, including environmental exposures, lifestyle and genetic susceptibility [157]. The first breakthrough in PD genetics came in 1997 when a missense mutation in the gene encoding α-synuclein (SNCA) was found to be asso-

ciated with the autosomal dominant PD phenotype [297]. Since then, several other genes associated with autosomal dominant PD, including coiled-coil helix coiled-coil helix domain 2 (CHCHD2) [175], homolog subfamily C member 13 (DNAJC13) [25], Eukaryotic translation initiation factor 4 gamma 1 (EIF4G1), HtrA serine peptidase 2 (HTRA2) [157], eucine-rich repeat kinase2 (LRRK2), ubiquitin c-terminal hydroxylase 1 (UCHL1), and vacuolar protein sorting 35 (VPS35) [25] were identified through several studies. Further, mutations in LRRK2 gene was found to be the most common monogenic PD in the Ashkenazi and Arab population [308]. Similarly, genes associated with autosomal recessive PD were also discovered through genetic studies. These include ATPase Cation Transporting 13A2 (ATP13A2), DNAJ/HSP40 homolog subfamily C member 6 (DNAJC6), F-box only protein 7 (FBX07), PTEN-induced putative kinase 1 (PINK1), parkin RBR E3 ubiquitin protein ligase (PRKN), synaptojanin 1 (SYNJ1) and vacuolar pro- tein sorting 13C (VPS13C) [66, 219]. These findings highlight the genetic heterogeneity of monogenic PD, which may explain the difference in clinicopathological features, dementia risk, ages of onset and the presence and extent of Lewy body pathology in PD cases.

Kelin and Westenberger (2012) estimated monogenic forms of PD to account for 30% of familial and 3%–5% of sporadic cases [157]. However, numerous genetic studies have highlighted other forms (not related to monogenic forms) of PD. Notably, mutations in the gene Glucosylceramidase Beta 1 (GBA), which encodes for glucocerebrosidase, an enzyme involved in metabolising glucosylceramide, are considered to be a major genetic risk factor for PD [74]. Further, impaired activity of the enzyme was shown to cause an accumulation of α-synuclein. Moreover, in a population-specific study, Velez-Pardo et al. (2019) found that age at the onset of PD was significantly earlier in GBA carriers than non-carriers [321]. As of today, several GWAS studies have successfully mapped several loci associated with sporadic PD. Nalls et al. (2014) used data from over 14,000 individuals (including PD patients and control) to identify risk loci for sporadic and familial PD[223]. Surprisingly, the loci were found in genomic regions containing the genes SNCA and LRRK2. In a similar study, Nalls et al. (2019) used data from over 1.5 million subjects to identify 90 genome-wide significant risk loci (38 novel risk variants) across 78 genomic regions. Other notable GWAS studies: PD progression GWAS (Iwaki et al. (2019);

Tan et al. (2020)) [132, 304]; and PD age at onset GWAS (Blauwendraat et al. (2019)) [24], highlight that a different set of genes may play a role in the age of onset and/or progression of PD once it is established. For example, PINK1, PRKN, SNCA, LRRK2 and Transmembrane Protein 175 (TMEM175) are genes associated with PD age at onset, while Solute Carrier Family 44 Member 1 (SLC44A1), GBA and Apolipoprotein E 4 (APOE4) were associated with PD disease progression [132, 304].

Based on the overwhelming evidence of shared clinical and pathological features of Lewy Body diseases, several studies were carried out to assess the genetic overlap between the diseases. Meeus et al. (2012) [213] performed a gene-based mutation analysis using genotype data from PDD and DLB cases (including familial and sporadic forms) and found several AD and PD-related genes to contribute to the genetic aetiology of PDD and DLB. For example, pathogenic missense mutations were detected in the AD genes in Presenilin 1 (PSEN1), Presenilin 2 (PSEN2), LRRK2 and icrotubule-associated protein tau (MAPT), and duplications in PD-related gene SNCA [213]. Similarly, Clark et al. (2009) [54] found GBA mutations to be associated with cortical Lewy body pathology and a less severe form of AD after adjusting for sex, age of death and APOE4 status, which is the strongest risk factor gene for AD [54]. Recently, a Lewy body dementia GWAS identified novel risk loci with a similar direction of effect in the genes GBA, TMEM175, Apolipoprotein E (APOE) and Bridging Integrator 1 (BIN1), identified in previous AD and PD GWAS [52]. Further, Guo et al. (2022) [107] used GWAS summary statistics from late-onset AD GWAS, PD GWAS and LBD GWAS and found positive genome-wide genetic correlations between the four neurological conditions. The strongest correlation was found between PD and Lewy Body dementias (PDD and DLB) (rg = 0.6352, se = 0.1880; P = 0.0007), while the weakest was between AD and PD (rg = 0.2136, se = 0.0860; P = 0.0130) [107]. Further, the authors carried out pathway enrichment analysis of 40 common genes identified through transcriptome-wide association analysis (TWAS) and found significant enrichment for GO term related to neurofibrillary tangle assembly. Genetic risk scores from these GWAS demonstrate that AD and PD share genetic risk factors and pathways (including lysosomal and endocytic related pathways) with Lewy Body dementias [107].

2.1.3 The influence of environment, diet and lifestyle factors

Although GWAS studies have found numerous genetic variations contributing to the risk and progression of Lewy Body diseases (with heritability estimates spanning from 20 to 30%) [151, 103, 106], environmental factors are still among the risk factors for neurological diseases. Recently, epidemiological studies have found several environmental factors, such as exposure to pesticides, cigarette and drug intake, abdominal obesity (body mass index (BMI)), caffeine and alcohol consumption, and physical activity, to be associated with PD risk [307, 217, 169]. For example, in 1976, a neurotoxin, 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) which kills dopaminergic neurons, was accidentally synthesised by a student who was trying to make synthetic heroin and which resulted in devastating parkinsonism in the predominantly young people exposed to this toxin [169]. Again, in 1983, in a peculiar but similar case involving young drug addicts, individuals who were exposed (injected intravenously) to the drug meperidine (MPTP is a potent analog of meperidine), developed parkinsonism [169]. Further investigation revealed that their parkinsonism was caused by the MPTP by-product of 1-methyl-4-phenyl-4-propionoxypiperidine (MPPP) synthesis. MPTP is a lipid-soluble neurotoxin that has the potential to cross the blood-brain barrier. Therefore, the drug is widely used to mimic PD symptoms in an experimental PD model [217, 310].

In an earlier study, Tanner et al. (2011) [307] applied a case-control study to explore the relationship between PD risk and certain pesticides. The findings showed a positive association between PD risk and exposure to a group of pesticides (including rotenone (OR=2.5; 95% CI, 1.3-4.7)) that inhibit mitochondrial complex I and other pesticides (including paraquat (OR=2.5; 95% CI, 1.4-4.7)) that cause oxidative stress. Interestingly, paraquat and rotenone (well-known pesticides used in animal models of PD) share a similar chemical structure with MPTP. Similarly, Xu et al. (2016), [346] performed a systemic search for studies on PDD risk factors and found a positive association between PDD risk and several non-genetic factors, including advanced age (odds ratio [OR] 1.07, 95% CI 1.03-1.13), male (OR 1.33, 95% CI 1.08-1.64), hallucination (OR 2.47, 95% CI 1.36-4.47), sleep behaviour disorder (OR 8.38, 95% CI 3.87-18.08) and smoking (RR 1.93, 95% CI 1.15-3.26). Interestingly, the authors found higher

education (RR 0.94, 95% CI 0.91-0.98) a protective factor for PDD. Further, other studies have highlighted that cognitive reserve may be protective against the progression of dementia [224]. However, more studies are needed to draw definitive conclusions. In 2020 the Lancet Commission report on dementia prevention identified 12 modifiable risk factors for dementia, including alcohol consumption, air pollution, diabetes, depression, obesity, traumatic brain injury (TBI), early education, hypertension, smoking, hearing loss, lack of physical activity and social isolation (which were estimated to account for 40% of worldwide dementia cases, although the prevalence of risk factors differ between countries) [187].

Given that epidemiological studies are susceptible to various biases [173], residual confounding, confounding by indication and reverse causation, their findings may not be reliable enough to establish causal relationships between environmental factors and PD risk. Therefore, researchers have employed Mendelian randomisation (MR), which uses genetic variants as instrumental variables to examine the causal effects of modifiable risk factors on disease outcomes [173]. For example, Heilbron et al., 2021 applied MR (using data from over 2 million PD cases and controls) to examine causal relationships and found smoking (OR 0.955; 95%confidence interval [CI] 0.921-0.991; p = 0.013) and higher BMI (OR 0.988, 95%CI 0.979-0.997; p = 0.008) to have a protective effect on PD; and higher daily alcohol intake (OR 1.125, 95%CI 1.025-1.235; p = 0.013) to increase the risk of PD[118]. Therefore, in order to dissect the pathogenesis of Lewy Body diseases, it is more beneficial to use integrative genomics (by combining multiple form of genomic analyses including RNA-Seq technology, integrative GWAS, MR [173]) approaches to assess the aetiology of the diseases.

2.1.4 Aim and objectives

In spite of the substantial progress made in the effort to describe the overlapping genetic, clinical and neuropathological features of LBDs, the biological basis and cellular drivers of pathology remain unclear. In the current study, Single-nucleus RNA-Sequencing was applied to anterior cingulate cortex samples derived from twenty-eight subjects to assess the extent of molecular and cellular heterogeneity in Lewy Body diseases (Figure 2.1). The anterior cingulate cortex was picked as a region of interest because previous studies have shown it to be one of the first cortical regions to be affected by Lewy body pathology and Lewy Body density in the region correlated with the development of dementia in PD-related studies [327, 131]. The objectives of the research reported in this Chapter were as follows:

- To generate source- single nucleus RNA-sequencing profiles across Lewy body diseases (PD, PDD, DLB).
- To assess cell-type specific transcriptional changes (namely excitatory neurons, inhibitory neurons, microglia, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, and vascular cells) in PD, PDD and DLB.
- To assess cell-type transcriptional changes associated with Lewy body dementia (PDD and DLB compared to control and PD brains).
- To identify disease-associated cell types and biological processes enriched in the set of genes differentially expressed.
- To test for heritability of enrichment for relevant phenotypes (e.g., PD, AD) in the differentially expressed genes across the disease states.

2.2 Methods

2.2.1 Sample and tissue sectioning, nuclei and RNA isolation

Professor Steve Gentleman and Dr Bension Tilley [310] performed sample selection. PD, PDD, DLB and neurologically "healthy" individuals were obtained from the Parkinson's UK Tissue Bank. In total 21 cases (n=7 per condition) and 7 control samples were matched for age and other demographic and pathologic factors where possible. Parkinsonism was classified based on the current official International Parkinson and Movement Disorder Societ (MDS) Clinical Diagnostic Criteria for Parkinson's disease (MDS-PD criteria)[248], while PDD was defined by the latest clinical diagnostic criteria for dementia associated with Parkinson's disease [81]; and DLB was defined using the criteria set by the fourth consensus report of the DLB Consortium [81].

Furthermore, PDD and DLB individuals were separated based on positive clinical features, including fluctuating cognition, visual hallucination and rapid eye movement sleep behaviour disorder (RBD); and the "1-year rule" used to differentiate between PDD and DLB. Routinely used diagnostic process, Braak α-synuclein and Tau, were used for pathological assessment of the tissue samples. In order to ensure dementia arose from α-synucleinopathy, individuals with severe AD pathology were filtered out using Braak tau stage 3 as the maximum threshold. PD individuals were selected based on a lack of evidence for any cognitive impairments, including memory impairment, executive dysfunction and visuo-spatial dysfunction in the clinical case notes. Neurologically "healthy" controls were selected based on a lack of neurological features alongside other pathological diagnoses (refer to Appendix A, Table 1 for sample demographics and clinical and pathological measures).

All lab-based sample processing steps, including tissue sectioning, were carried out by Dr Amy Smith in the following steps: (i) 80 µm tissue block of cortical grey matter was sectioned from the anterior cingulate cortex, and adjacent sections were used for nuclei isolation; (ii) Buffer solutions (e.g., Homogenisation Buffer) were prepared using a combination of protocols from

10X Genomics and Krishnaswami et al. (2016) protocol; (iii) Tissue sections were suspended in 800 µL HB, homogenised, filtered through a BD Falcone tube and centrifuged at 1000g for 8 minutes. Further clean-up of nuclei was performed as per Krishnaswami et al. (2016) [160] with some adjustments including using 29% iodixanol solution at 13,000 g for centrifugation of the layered nuclei and adjusting the time to 20 minutes instead of 40 minutes suggested by the authors; (iii) the supernatant was removed, nuclei pellet washed with PBS bugger, filtered, centrifuged and washed again (all specific steps including time take for each step can be found in Krishnaswami et al. (2016) protocol); (iv) finally, nuclei were counted using Acridine orange dye routinely used to stain nuclei. RNA isolation was performed by BioXpedia using the following protocol: (i) samples were lysed and RNA extracted using QIAzol and Qiagen's RNeasy 96 Kit, respectively, (following manufacturer instructions). Quality of samples were evaluated by measuring their RNA integrity number (RIN), obtained using Agilent 4200 Tapestation (Agilent). Only samples with RIN value of 4.2 or above were selected for downstream analysis.

2.2.2 Single-nucleus RNA sequencing data generation and processing

snRNA-Seq data was generated using 10X Genomics Chromium Single Cell 3' Reagent Kit (v2 chemistry) following 10X protocols. Briefly, (i) approximately 10,000 nuclei were targeted, followed by amplification process (8 cycles of cDNA and PCR indexing (14 cycles); (ii) Qubit dsDNA HS Assay Kit (ThermoFisher) was used to measure cDNA concentration, while Bioanalyzer High-Sensitivity DNA Kit (Agilent) was used to assess quality and length of libraries; (iii) finally, samples were pooled and sequenced on 8 lanes of Illumina Hi-Seq 4000 high-throughput sequencer.

2.2.3 Mapping and post-alignment quality control

Generation of sequencing reads was carried out using Cell Ranger (v 3.0.2) protocol in the following steps: (i) raw base files were converted to FASTQ files; (ii) reads were mapped to exonic and intronic regions of the reference human genome version GRCh38, and gene

annotations were obtained from Ensembl v93; (iii) finally, gene barcode (unique molecular identifier (UMI)) matrices were generated for each sample. Following alignment to a reference genome, each sample was assessed for the following RNA-Sequencing metrics: (i) percentages of mapped reads; (ii) number of genes detected; (iii) average UMI count. All samples passed the threshold set for each metric (Figure 1). Quality control of reads was carried out in the following steps: (i) reads from "empty" droplets, a droplet which contains ambient RNA, were removed using **EmptyDrops** algorithm from the R package `DropletUtils` [193]. The algorithm removes nuclei droplets containing low amounts of RNA by applying the Dirichlet-multinomial model of UMI count sampling in order to distinguish true nuclei from "empty" nuclei based on their deviation from the calculated profile of an ambient RNA pool (a pool of low amounts of RNA containing nuclei in datasets; and the parameter for UMI count (a measure of RNA content) was threshold at 300, a value based on in-house calculations); (ii) `CreateSeuratObject()` from the R package `Seurat` (v 3.0) [41] was used to generate an R object for each sample; (iii) for each sample, genes containing greater than 5% mitochondrial reads (mitochondrial genes with the prefix MT-) were removed; and genes detected in less than 5 nuclei were removed.

2.2.4 Normalisation, clustering and doublet detection

Once low-quality reads and genes were removed: (i) each sample was normalised using the Seurat's function `LogNormalize()`, in which counts for each nucleus are divided by the total counts for that nucleus and multiplied by a scale factor (set at 10,000 based on targeted number of nuclei) and natural-log transformed; (ii) each sample was then clustered using the function `FindClusters()` with the parameters 30 for principal components (principal components (PC)) selection and 2 for resolutions and visualised using a non-linear dimensionality reduction algorithm known as UMAP; (iii) finally, potential doublets or multiplets were removed using a doublet prediction function `doubletFinder_v3()`, as implemented in the R package `DoubletFinder`[209]. Briefly, the algorithm removes technical artefacts from each cluster (per sample) by first computing the distance between each nucleus to artificially generated doublets (which are generated by calculating the average expression profile of randomly selected nuclei

pairs), and if a nucleus has a higher proportion of artificial nearest neighbours, it is labelled as a "doublet" and removed from the cluster.

2.2.5 Joint graph generation and cell-type identification

Once the doublets or multiplets were removed, the Seurat function `FindAllMarkers()` was used to identify differentially expressed genes in each cluster compared to all other clusters. Each cluster was then assigned a cell-type by first calculating (using Fisher's exact test) the overlap between each cluster marker genes and a cell-type marker gene list derived from two human single-cell datasets [332]; and assigning a cell-type to a cluster based on the lowest p-value. A joint graph of all nuclei across all individuals was generated using **Conos** [**19**] pipeline. Briefly, pair-wise comparisons were performed across the datasets, followed by inter-sample edges calculations, combined with intra-sample edges to generate a joint graph; finally, cell-type labels were propagated using a randomly selected pre-labelled nuclei vector.

2.2.6 Generation of condition-specific cell-type specificity matrices

For each group (control, PD, PDD and DLB, separately), a gene cell-type specificity matrix was generated using the `generate.celltype.data()` function from the R package `Expression Weighted Celltype Enrichment (EWCE)` [285], a method which statistically assesses whether a specific cell type exhibits a higher expression of a set of genes than what would be expected by random chance. The matrices contain values from 0 to 1, with 0 value implying a gene is not expressed in a cell-type, while 1 implies a gene is expressed only in that cell-type. These values represent the proportion of the total expression of a specific gene in that cell-type.

2.2.7 Differential expression analyses

Differential expression analysis was carried out to compare transcriptional differences and similarities between the four groups. Briefly, separate matrices were generated for the seven cell-types identified in the dataset, and DEA was performed using the tool MAST [89] (refer to the method section in Chapter 1.1.2 Downstream data analysis, for a brief description of MAST).

To explore gene expression differences across all cell-types between the three groups, after adjusting for sex, with age at death, post-mortem interval (PMI), CDR (which acts as a substitution for intrinsic and extrinsic factors that could influence gene expression) and individual identification number, the following comparisons were taken: 1) Comparisons between disease states (PD,PDD and DLB) and control (case-control); 2) Comparisons between disease states (case-case comparisons; a full list of comparisons can be found on Table 2.1). Genes with FDR < 0.05 and absolute fold-change > 1.5 were considered significant (differentially expressed).

2.2.8 Functional enrichment analysis

Functional term enrichment analyses were run for cell-type-specific differentially expressed gene sets from each pairwise comparison using the R package `WebGestaltR` (version 0.4.4) [182]. The analyses were carried out using (i) GO biological process terms and (ii) pathways associated with increased PD-risk in a large-scale pathway-specific polygenic risk analysis and rare variants analysis. The default values for `WebGestaltR` parameters were used, and false discovery rate (FDR) were adjusted using Benjamin and Hochberg (BH) FDR correction method. Pathways with FDR < 0.05 were considered significantly enriched. Furthermore, GO-derived pathways were reduced for redundant terms using the R-based functions from `rutils` `https://github.com/RHReynolds/rutils`, which is a combination of two R packages, `GOSemSim` [352] and `rrvgo` `https://ssayols.github.io/rrvgo`. Briefly, GO terms with less than 20 genes or more than 2000 genes are removed and semantic similarity scores are calculated for each enriched GO term using the function `mgoSim()` from the `GOSemSim` package. The function calculates the scores using a hierarchical clustering approach (complete linkage). The hierarchical tree was then cut at a threshold of 0.7. Next, a distance matrix is calculated from the semantic similarity scores using the function `reduceSimMatrix()` from the `rrvgo` package. Finally, the parent term with the highest semantic similarity score was used to represent each group of child terms.

2.2.9 H-MAGMA and LDSC

Genetic enrichment of association analyses were performed using two methods:Hi-C-coupled MAGMA (H-MAGMA) and sLDSC. Two different methods with different assumptions were used to increase the credibility of the genetic associations.

Both methods were run with two sets of gene lists: (i) DE gene sets across all comparisons and (ii) Cell-type specificity matrices (per group) calculated using the R package `EWCE`, which uses expression data to impute specificity values from 0 to 1. The annotations (gene lists) were run with 3 GWAS, including Parkinson's disease (PD)[222], Parkinson's disease Age of Onset (PD AOO)[24] and Alzheimer's disease (AD)[134]. For all the analyses, p-values were FDR calculated for the number of cell-types tested. Cell-type specific differentially expressed genes were tested to see they are genetically associated with a disease state (PD, PD-AOO, AD). H-MAGMA, an extension of MAGMA, uses fetal and adult brain Hi-C to compute chromatin interactions to exons and promoters in order to assign intronic or promoter (2kbp upstream of transcription start site) and exonic SNPs to their cognate genes based on their genomic location. The parameters used to compute association statistics were 10kb downstream and 35kb upstream. Similar parameters were applied to test for top 10% most specific gene sets in each group per cell-type.

stratified LDSC (v 1.0.0)(sLDSC)[90] was run by Dr Regina Reynolds. sLDSC was used to test for genetic association between genetic risk of PD, PD-AOO, and AD and cell-type specific DEGs or top 10% most specific gene sets per group. In order to capture regulatory elements, the window coordinates were extended by 100kb upstream and downstream of their transcription start and end site. Furthermore, all SNPs were a minor allele frequency of 0.05 were used to define a SNP using binary terms 0 (absent) and 1 (present). Next, the annotations were added to an existing baseline model of 53 annotations (provided by Finucane et al. (2015) (v 1.2)[90]) to generate new annotation files. Thereafter, HapMap Project Phase 3 (HapMap3) [60] SNPs were used for regression and 1000 Genomes Project Phase 3 [59] European population SNPs were used for LD reference panels. Due to their complexity and distance, the major

histocompatibility complex regions were excluded from the analyses. Finally, using coefficient z-scores from the association statistics one tailed p-values were imputed and corrected for multiple testing (FDR). Those values with a positive contribution to a disease heritability were reported.

2.2.10 Sub-clustering analysis

In order to analyse level 2 (for example neuronal sub-types) cell-type specific clustering was performed for neurons (since they were the cell-type mainly highlighted by DEA and pathway enrichment analysis). Briefly, new matrices were generated for neuronal cell-type based on previous classification (refer to Chapter 2.2.5). For each matrix, a new Seurat object was generated and the following analyses were performed: (i) normalisation; (ii) clustering: (iii) cell-type identification. To classify the neuronal nuclei, a gene list curated by combining the top 10% most cell-type specific marker genes from Skene et al., 2018 [284] was used.

2.2.11 Code and Data availability

Code used to process and analyse single-nucleus RNA-sequencing data and to generate H-MAGMA outputs is available at: https://github.com/rahfel/snRNAseqProcessingSteps . snRNA-Seq data can be accessed through the Gene Expression Omnibus (accession ID: GSE178146).

2.3 Results

2.3.1 Study design

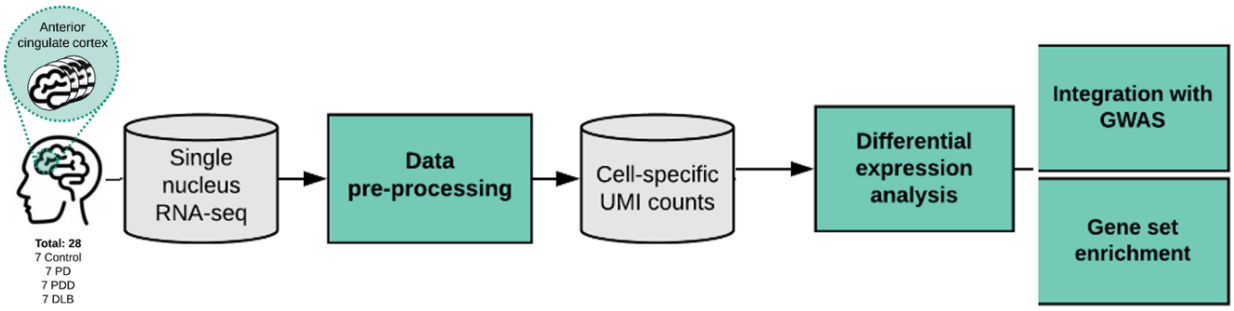


Figure 2.1: **Overview of approach.** Single-nucleus RNA-Sequencing data generated using 28 subjects. Abbreviation: GWAS, Genome-wide Association Studies. UMI, Unique molecular identifiers; PD, Parkinson's Disease; PDD Parkinson's Disease with Dementia; DLB, Dementia with Lewy Bodies.

Anterior cingulate cortex tissue samples were taken from a cohort of 7 non-neurological controls, 7 Parkinson's Disease without cognitive impairment (PD), 7 PDD and 7 DLB (Figure 2.1). Thereafter, snRNA-Seq data was used to generate cell-type-specific differential gene expression profiles.

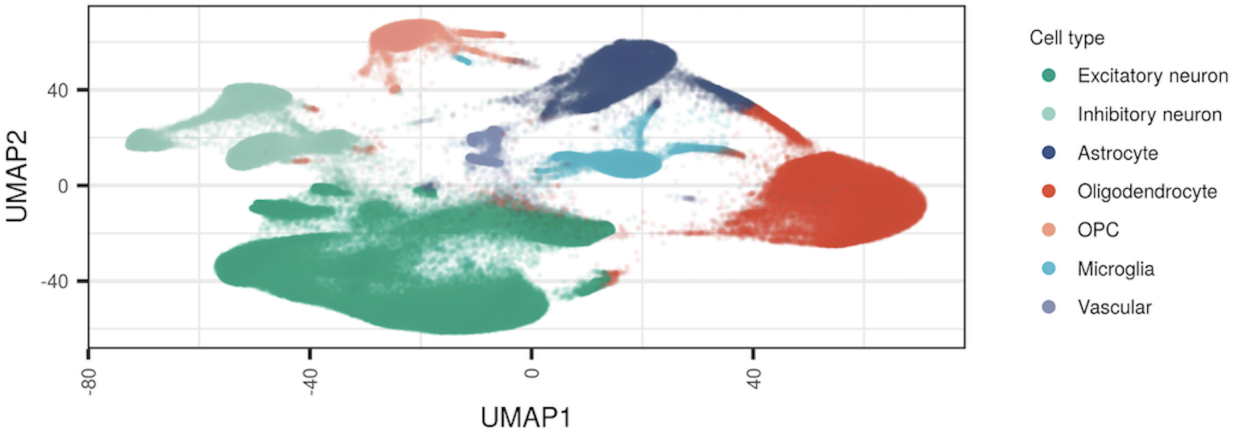
2.3.2 Data generation and pre-processing

The tissue samples were taken from the anterior cingulate cortex, a region of the brain in which Lewy body densities correlate with Lewy Body Diseases (PD, PDD and DLB) severity. The subjects were matched for demographic and pathologic factors. However, there were significant differences in disease duration and the proportions of sexes between the group (which is also reflective of male bias in PDD and DLB previously described in the literature [225]) the ratio in each group was: 6 male / 1 female in control, 2 male / 5 female in PD, 5 male / 7 female in PDD and 7 male / 0 female in DLB. The mean disease duration in years for each group was: PD = 12, PDD = 11, and DLB = 6. The fact that DLB subjects had the lowest recorded disease

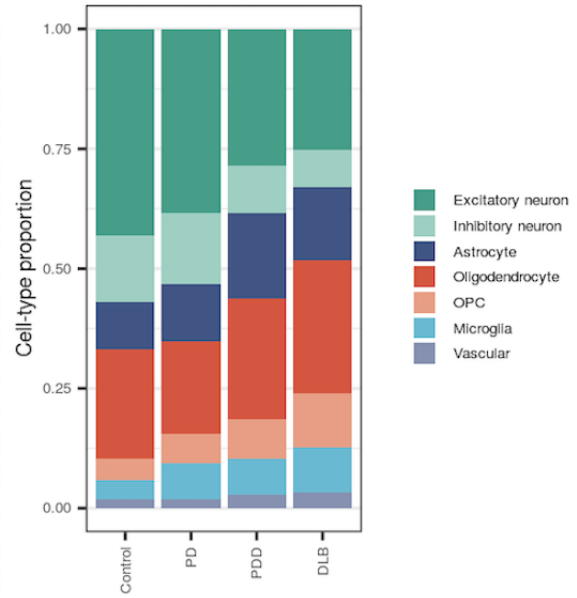
duration before death may reflect parkinsonism appearing several years before the development of cognitive impairment in subjects with PDD. Each group's mean age at death was: control = 75, PD = 83, PDD = 78 and DLB = 73. Median Braak tau stage was: controls = 1, PD= 2, PDD =2 and DLB = 2 (we considered a tau Braak state less than 3 to be indicative of no significant tau pathology). The median Braak α-synuclein stage was: control = 0, PD = 6, PDD = 6 and DLB = 6. Alafuzoff α-synuclein staging was used to stain the density of α-synuclein in each brain tissue sample. The density was observed to increase in the order of PD > PDD > DLB. (The subjects' full clinical and demographic data are in Appendix A, Table 1). A total of 205,948 droplet-based single-nucleus were obtained from the samples after initial quality control (QC) assessments. The mean values for the QC metrics were: (i) an average of 89% of reads were mapped to the genome; (ii) an average of 1,398 genes were detected per nucleus; (iii) an average of 7,355 nuclei were detected per sample; (iv) an average 2,358 UMI counts were detected per nucleus (average of 98% valid barcodes). (Full QC metrics data can be found in Appendix A, Table 2)

2.3.3 Joint graph generation

A joint graph of all nuclei across all subjects was generated using `Conos` algorithm which computes a graph by estimating multiple inter-sample mappings (Figure 2.2). Each cluster was assigned a cell-type by assessing the overlap between Differentially Expressed Genes (DEGs) sets in each cluster and a list of marker genes derived from Wang et al. (2018) [332] in which a list of cell-type marker genes was generated by combining two human single-cell datasets. Fisher's exact test was used to carry out the significant overlap ($p < 0.01$). In total, across the full sample set, 75,826 excitatory neurons, 46,662 oligodendrocytes, 26,467 inhibitory neurons, 25,726 astrocytes, 12,497 oligodendrocyte precursors (OPCs), 13,788 microglia, and 4,532 vascular cells (including endothelial cells and pericytes) were identified. The proportions of cell types across each group were visualised using the R package `ggplot2` (Figure 2.2). Full number of nuclei per individual can be found in Appendix A, Table 3.



(a) **Joint graph of all nuclei derived from all individuals**



(b) **Cell-type proportions per disease group.**

Figure 2.2: **Cell-type proportions.**Proportion of Excitatory neurons, Inhibitory neurons, Microglia, Astrocytes, Oligodendrocytes, Oligodendrocyte precursor cells and Vascular (a mixture of pericytes and endothelial nuclei) derived from 27 subjects (n = 7 per disease group). Abbreviations: OPC, Oligodendrocyte precursor cells

Comparison type	Comparisons	Total number of subjects	Cases	Controls
Case-Control	Control vs DLB	14	7	7
	Control vs PDD	14	7	7
	Control vs PD	14	7	7
Case-Case	PDD vs DLB	14	14	0
	PD vs DLB	14	14	0
	PD versus PDD	14	14	0

Table 2.1: **Number of subjects per group**

2.3.4 Cell-type specific differential gene expression analysis

Differential expression (DE) analysis was carried out to explore the transcriptional changes associated with the disease state versus controls (case-control comparison) and between conditions (case-case comparison). The analyses were performed separately for each cell-type in a pair-wise manner (Table 2.1).

After adjustment for the covariates sex, age, PMI and CDR- defined as the fraction of genes that are detectable in each cell, 9,242 genes were found differentially expressed ($|\log2(\text{fold change})| > \log2(1.5)$, FDR < 0.05) across all six pairwise comparisons across all seven cell-types (Figure 2.3a). Most of the differentially expressed genes (DEGs) overlapped across all comparisons, suggesting that most variable genes were similar across disease groups and reflective of the LBDs continuum hypothesis mentioned in previous studies. Further, the significant overlap of DEGs suggest that gene expression changes may not be entirely driven by the differences in cell-type proportions in the groups. Out of all the DEGs, approximately 36% were found to be uniquely (only found in one comparison and in one cell-type) differentially expressed in cell-type specific pairwise comparisons (across the six comparisons), in line with the view that gene expression changes exist between disease groups (Figure 2.3b).

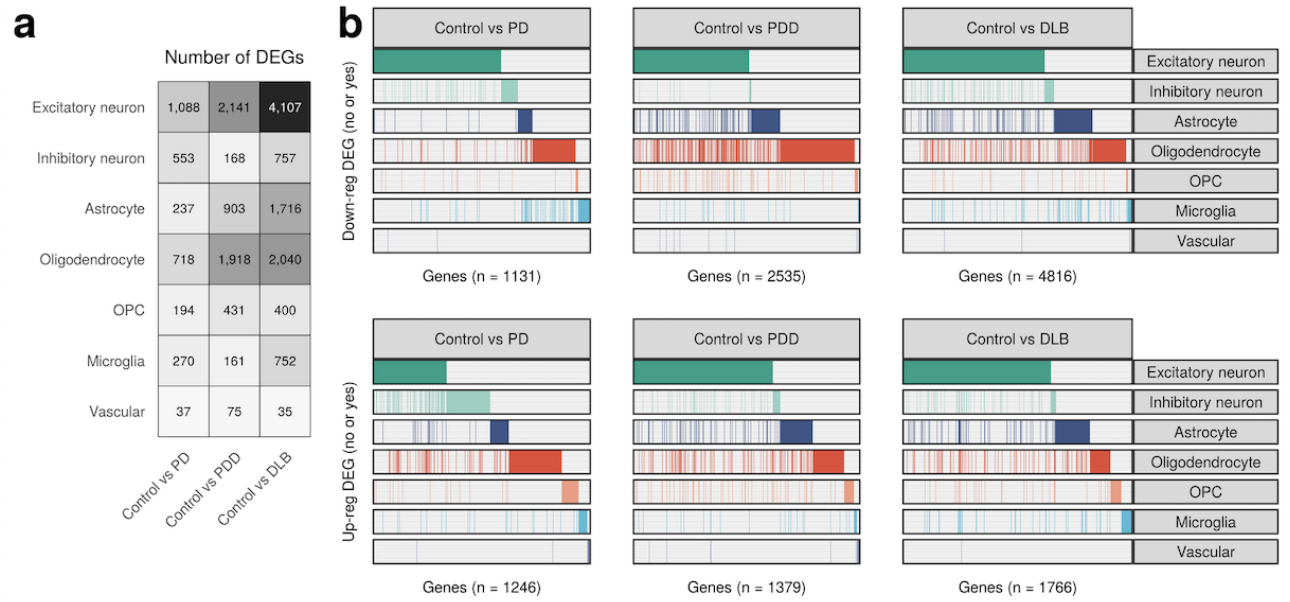


Figure 2.3: **Gene expression changes and perturbed pathways across cell-types and disease state**. a Total number of differentially expressed gene (DEGs) across each pair-wise comparison ((|log2(fold change) | >log2(1.5), FDR < 0.05). b Binary plot (yes or no) indicating with bars whether a specific gene (column) is down-regulated (upper panel) or up-regulated (lower panel) in a given cell type (rows). Number of DE genes in each comparison indicated on the x-axis. All results related to cell-type specific differential expression analysis and functional pathway enrichments across all comparisons can be found in Appendix A, Table 4 and Table 5, respectively. This figure was created for Feleke et al. (2021), which was generated during the course of this work and subsequently published [87]

Focusing on PD versus control comparisons, excitatory neurons had the largest number of DEGs (423 genes were up-regulated (have a higher expression in PD) and 665 down-regulated), followed by oligodendrocytes (436 gene up-regulated and 282 down-regulated) (Appendix A, Table 4 for a list of DEGs across all comparisons). For PDD versus control comparisons, there were more DEGs found in excitatory neurons (848 up-regulated and 1,293 down-regulated) as well as oligodendrocytes (392 up-regulated and 1,526 down-regulated). Expectedly, the total number of DEGs were higher for the DLB versus control comparison than any other comparison carried out: for excitatory neurons there were 1,136 genes were up-regulated and 2,971 down-regulated genes (nearly 4.5-fold found in PD versus control comparison); and for oligodendrocytes 353 gene up-regulated and 1,687 down-regulated genes.

Interestingly, among the genes differentially expressed (up-regulated in excitatory neurons in cases) across all three diseases (i.e., PD, PDD and DLB) versus control comparison, were genes involved in synaptic plasticity, signalling and memory formation including Calcium/Calmodulin Dependent Protein Kinase II Delta (CAMK2D), Calcium/Calmodulin De- pendent Protein Kinase II Beta (CAMK2B), G protein-activated inward rectifier potassium chan- nel 2 (KCNJ6), cyclic nucleotide phosphodiesterase 1 (PDE1A) and Tissue inhibitor of metal- loproteinases 2 (TIMP2). Furthermore, genes that were down-regulated in the excitatory neurons in the disease groups compared to control groups were genes involved in neurotransmitter pathways, including PD-associated genes calcium voltage-gated channel subunit alpha1 A (CACNA1A), calcium voltage-gated channel auxiliary subunit gamma 8 (CACNG8) and discs large MAGUK scaffold protein 2 (DLG2). These observations align with the notion that α-synuclein aggregates-related synaptic dysfunctions are associated with cognitive impairment.

2.3.5 Differentially expressed genes distinguish Lewy body dementias from Parkinson's disease

Looking at the comparison between disease and control groups separately from disease against disease comparisons, three main themes emerged (Figure 2.3):

1. There was widespread gene dysregulation in the glia and the neurons.

2. Most of the DEGs detected in the disease groups, compared to controls, were only DE in one cell-type. For example, focusing on PD versus Control comparison, out of the 1,246 up-regulated genes, 945 (76%) were specific to a single cell-type, while for down-regulated genes, 893 genes out of 1,131 (79%) were cell-type specific. Similar observations were made for PDD, with 74% for up-regulated genes and 66% for down-regulated genes; and for DLB, 74% and 67%, respectively.

3. The gene expression profiles of PDD and DLB were very similar, in keeping with the overlapping clinical features and pathology between these two conditions (Figure 2.3a). Notably, there were fewer transcriptional similarities observed between PD and PDD (Figure 2.3a), again consistent with cortical involvement in patients with PDD compared to those without dementia.

Variable expression of SNCA across disease groups

Next, differential expression of genes derived from a recently published study of mutations implicated in PD susceptibility [25, 16], including a well-characterised gene SNCA, were explored. Of the 21 genes considered, 13 were differentially expressed in at least one major cell-type in at least one disease versus control comparison (Figure 2.4a). For example, excitatory neurons, inhibitory neurons, astrocytes and oligodendrocytes all showed significant up-regulation of SNCA in PD cases when compared with controls (fold change: 0.64–1.30; FDR: 2.6×10^{-7} – 7.2×10^{-157}, Figure 2.4a).

Focusing on SNCA expression, the following observations were made (Figure 2.4b): (i) SNCA expression was up-regulated in PD relative to control samples across four cell-types (excitatory neurons, inhibitory neurons, astrocytes and oligodendrocytes); (ii) there was a decrease in the proportion of SNCA- expressing nuclei in PDD; (iii) there was a similar expression range of SNCA in control and PD; (iv) there was an absence of a neuronal subpopulation expressing higher levels of SNCA (top 10% highest expressing nuclei; Figure 2.4) in Lewy body dementias, as compared with PD and control group. The absence of nuclei with high SNCA expression in PDD and DLB may highlight a selective vulnerability of subpopulations of excitatory neurons

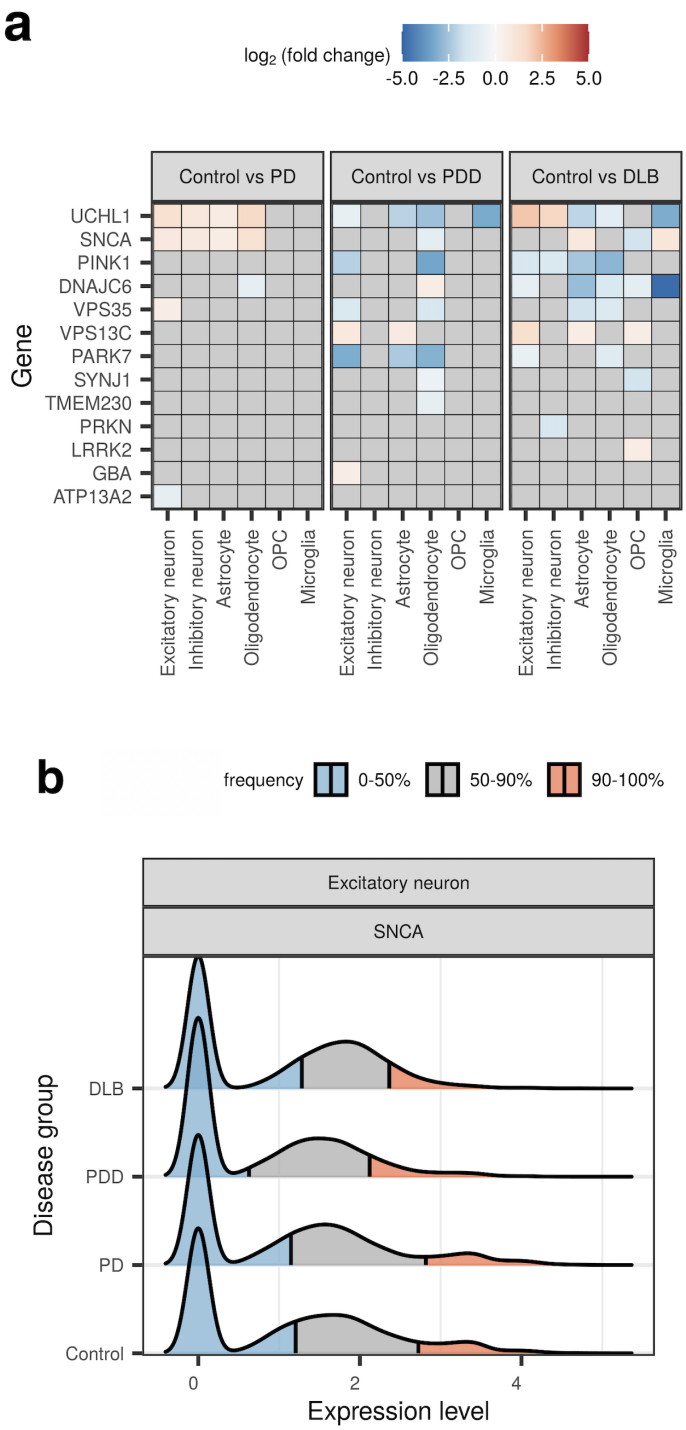


Figure 2.4: **Cell-type-specific alterations of PD-associated genes and pathways across disease versus control comparisons**. a Differential expression analysis of genes genetically linked to PD across cell-types and pairwise comparisons of disease groups with the control group. Non-significant genes (FDR > 0.05) are coloured grey, all other tiles are significantly expressed genes (FDR < 0.05). b SNCA expression distribution (in excitatory neurons) represented on ridgeline plot. Distributions have been split into 3 cumulative quantiles (0–50%, 50–90% and 90–100%) of the nuclei in each disease group. (Feleke et al., 2021)

which express high SNCA levels to neurodegeneration. Taken together, these observations support the notion that while there are transcriptional similarities between the Lewy Body diseases (based on the overlapping DEGs identified), PD has a distinct transcriptional profile from the other two Lewy bodies, likely relating to the absence of Lewy body pathology in the cortex of PD patients with solely motor features. In contrast, the Lewy body dementias PDD and DLB showed high levels of transcriptional concordance.

2.3.6 Identifying distinct and common molecular pathways in Lewy Body diseases

Functional enrichment analyses were performed on cell-type specific differentially expressed genes to explore their potential biological implications (Appendix A, Table 4 for terms identified using gene sets from case-case comparisons). Focusing on PD, PDD and DLB versus control comparisons, a total of 235 and 312 GO terms were enriched (significant at FDR < 0.05) for up and down-regulated DE gene sets, respectively. The significant pathways were reduced based on semantic similarity (carried out using R package `rutilis`). In total, 57 pathways for up-regulated and 68 pathways for down-regulated were enriched after adjustment for multiple testing (FDR < 0.05) (Figure 2.5).Focusing on the child terms which were significant using up-regulated DEG sets (including redundant terms), 106 out of 235 GO terms were uniquely enriched for a specific cell-type. For example, for DLB versus controls, GO terms including calcium ion transmembrane transport (FDR =0.037), calcium-mediated signalling (FDR = 0.023) and chemical synaptic transmission, postsynaptic (FDR $= 1.48 \times 10^{-07}$) were perturbed only in excitatory neurons. A full list of all significant terms for each comparison including child terms are found in Appendix A, Table 5. Furthermore, the highest number of enriched pathways were observed in those Lewy body diseases with cortical involvement (PDD and DLB), which is also reflected when comparing PD with PDD/DLB (Figure 2.6).

As well as analysing functional enrichment using GO categories, a set of curated PD-associated pathways were used from Bandres-Ciga et al. (2020) [16], in which the authors had prioritised 46 pathways for PD by using large-scale gene-set specific Polygenic risk score (PRS) based

Figure 2.5: **Pathways significantly enriched for up-regulated genes.**(Up-regualted in cases in comparison to control). Significance at FDR < 0.05. Abbreviations: GTP, guanosine triphosphate

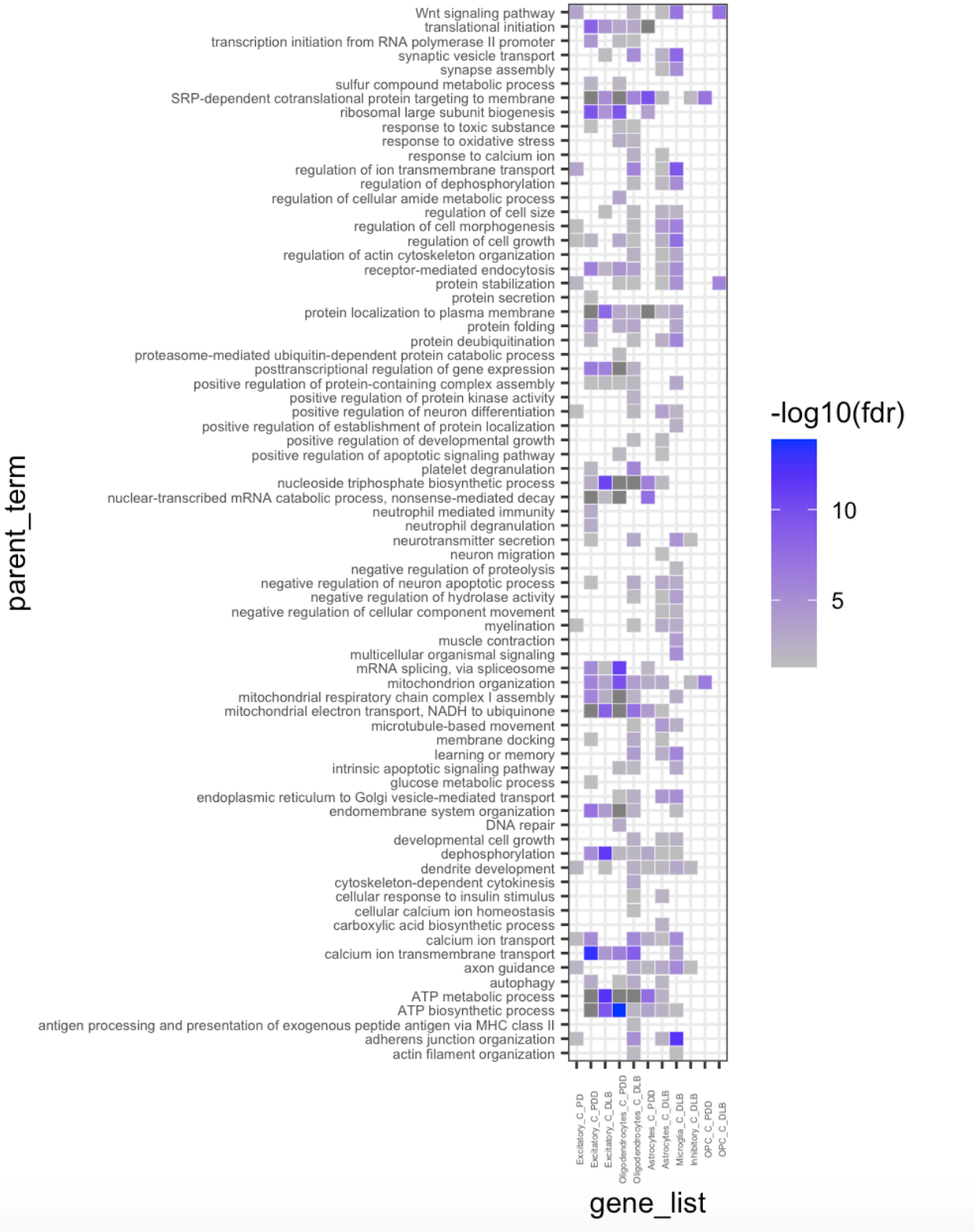


Figure 2.6: **Pathways significantly enriched for down-regulated genes.**(Down-regualted in cases in comparison to control).Significance at FDR < 0.05. Abbreviations: ERBB, epidermal growth factor receptor family; ATP, Adenosine triphosphate ; NADH, nicotinamide adenine dinucleotide (NAD) and hydrogen (H); SRP, signal-recognition particle

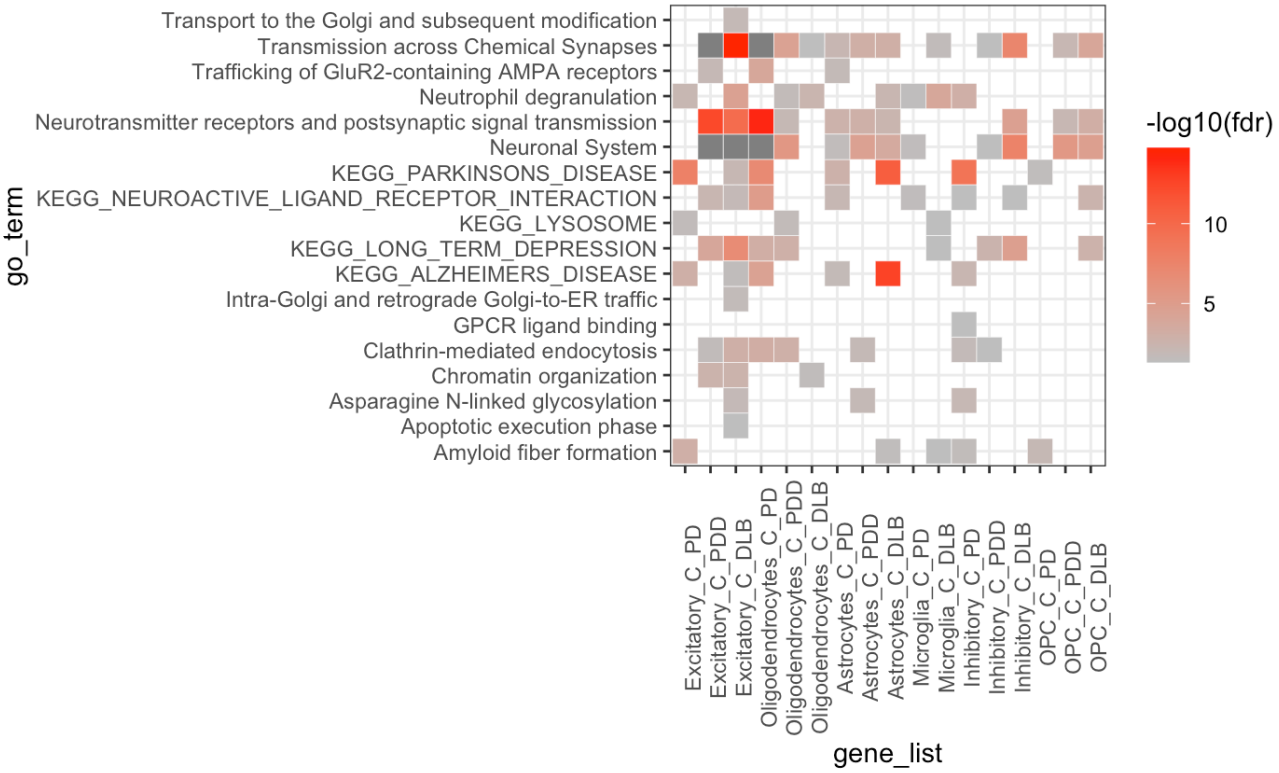


Figure 2.7: **Genetically PD-linked pathways significantly enriched for up-regulated genes.**(Up-regualted in cases in comparison to control). Significance at FDR < 0.05. Abbreviations: AMPA, α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; GLUR2, Glutamate Receptor 2; ER, Endoplasmic Reticulum; GPCR, G protein-coupled receptor; KEGG, Kyoto Encyclopedia of Genes and Genomes

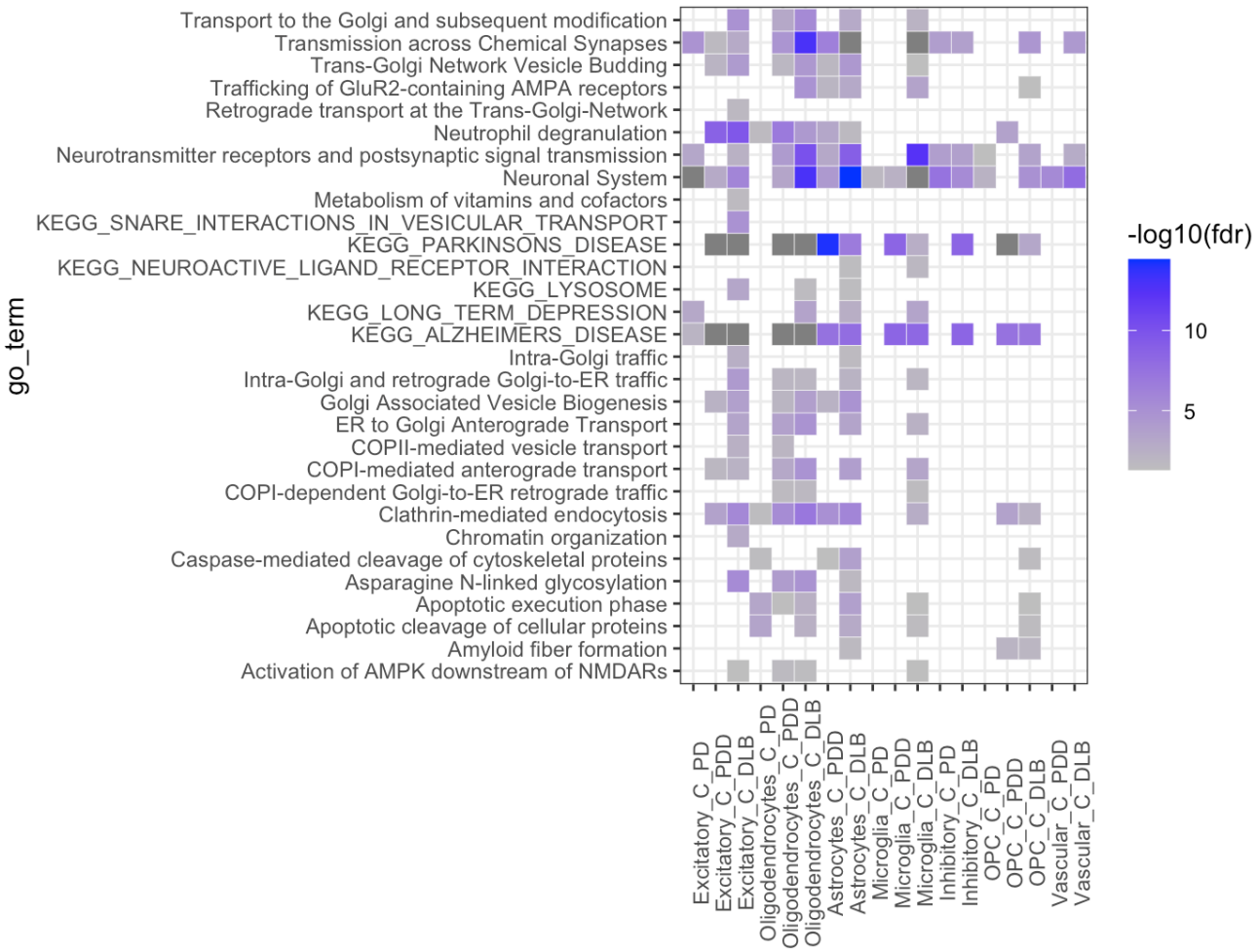


Figure 2.8: **Genetically PD-linked pathways significantly enriched for down-regulated genes.**(Down-regualted in cases in comparison to control). Significance at FDR < 0.05. Abbreviations: COPII, Coat Protein Complex II; NMDARs: Stands for N-Methyl-D-aspartate receptors

assessment of 2,199 gene sets to evaluate the role of common variation on PD risk[16]. Some of these pathways include terms related to neuronal transmission-related pathways, metabolism, programmed cell death and vesicle-mediated transport. In general, the number of perturbed pathways increased with increasing cortical involvement (i.e., PD -> PDD -> DLB) While most (17 of 31) pathways were enriched among up-and down-regulated genes, 13 pathways were uniquely enriched among down-regulated genes (Figure 2.7) and (Figure 2.8).(Appendix A, Table 6 for a full list of significant terms).

2.3.7 Heritability enrichment analysis

A major drawback of differential gene expression analysis, even at a single-cell level, is that transcriptional changes between conditions may arise both as a cause and a consequence of the disease. To try to identify transcriptional changes with a potential proximal relationship to disease, differentially expressed gene sets can be tested for enrichment of genetic variants that are associated with disease susceptibility. Here, two approaches with different assumptions and algorithms were used: (i) Stratified LD score regression (sLDSC)- a statistical method which tests the contribution of a genomic annotation to a trait heritability. In the current study, enrichment (the proportion of heritability explained by SNPs in the annotation divided by the proportion of SNPs in the annotation) was used to evaluate the significance of the association; (ii) MAGMA- a method which tests the collective association of all SNPs in a given gene with a trait while accounting for Linkage Disequilibrium between SNPs. In this study, H-MAGMA was used, which is an extension of MAGMA which uses chromatin interaction profiles from human brain tissue to assign intergenic and intronic SNPs to their nearest genes (refer to Method Section for full description of the methods).

Since co-existent AD pathology (for example, hyperphosphorylated tau and amyloid-β plaques) is a common feature in the Lewy body dementias (PDD and DLB), in addition to GWAS studies for PD age of onset [24] and PD-susceptibility [222], included GWAS summary statistics derived from a recent AD [134] GWAS was also used. For genetic enrichment analysis we considered cell-type-specific DE genes significant at FDR < 0.05 and $|\log2(\text{fold change})| > 1.5$), which

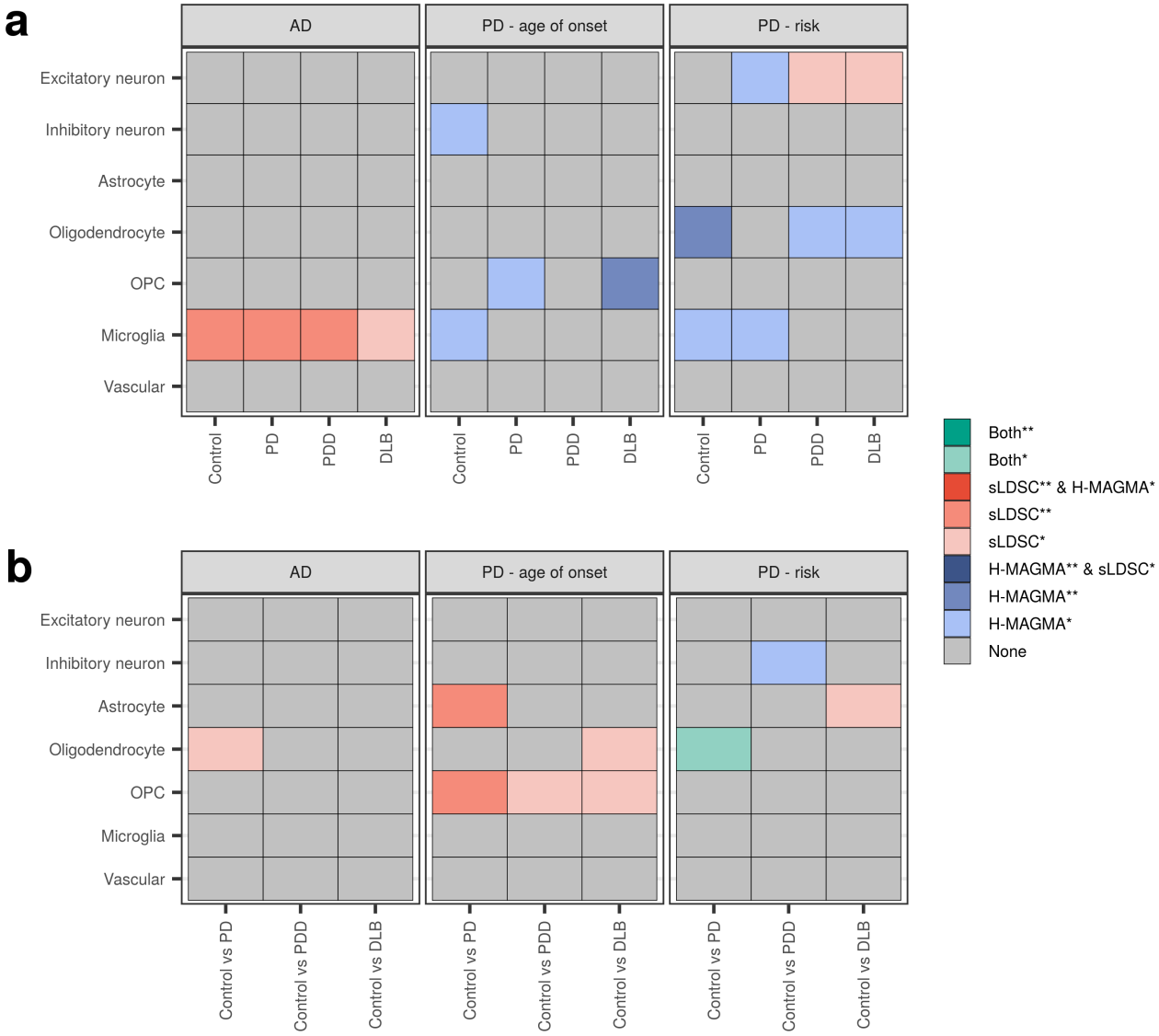


Figure 2.9: **Genetic association analysis using Hi–C-coupled MAGMA and stratified LD score regression.** a) association test using cell-type-specific differentially expressed genes in disease comparisons with controls (b) association test using top 10% most cell-type-specific genes in each disease group. Significance level are indicated by * for nominal significance (unadjusted p-value < 0.05) and ** for significant association (FDR-corrected p-value < 0.05) after adjusting for the number of cell-types tested.

were further split based on their direction of effect (Figure 2.9).

In summary, the concordance between the two methods was low; therefore, results obtained using both methods were reported separately in Appendix A, Tables 7 and 8. Using sLDSC, there was a significant association between up-regulated genes in OPCs and dysregulated (up and down) genes in astrocytes derived from PD versus control comparisons ($FDR_{LDSC} = 0.0076$ in OPC and $FDR_{LDSC} = 0.0085$ in astrocytes) and PD risk (Figure 2.9). Similarly, there was a significant association between up-regulated genes in excitatory neurons from comparison of PDD with control and PD genetic risk ($FDR_{LDSC} = 0.0040$). Furthermore, using HMAGMA there was a significant association between down-regulated genes in vascular and up-regulated genes in oligodendrocytes from PD-control comparison and PD genetic risk ($FDR_{HMAGMA} = 0.0082$ in vascular; $FDR_{HMAGMA} = 0.013$ in oligodendrocytes) (Figure 2.9; Full result table available in Appendix A, Table 7 and 8).

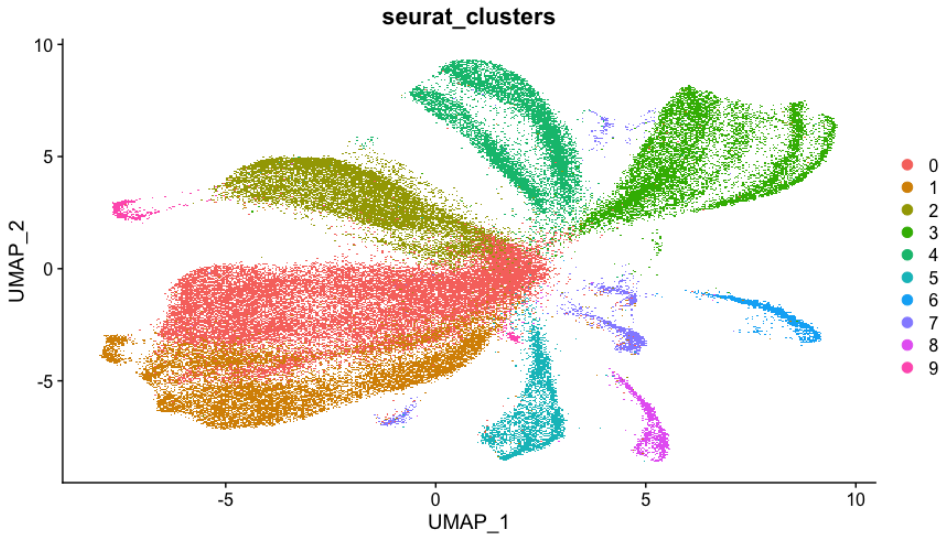
Neuronal sub-type preferential vulnerability in DLB

The above analyses revealed significant differential expression in all three diseases (PD, PDD, DLB) compared to controls in neurons, microglia and oligodendrocytes. Among these cell types, neuronal differential expression was enriched for functions relevant to the development of PD and genetic risk to PD, suggesting a possible aetiological relationship to PD.

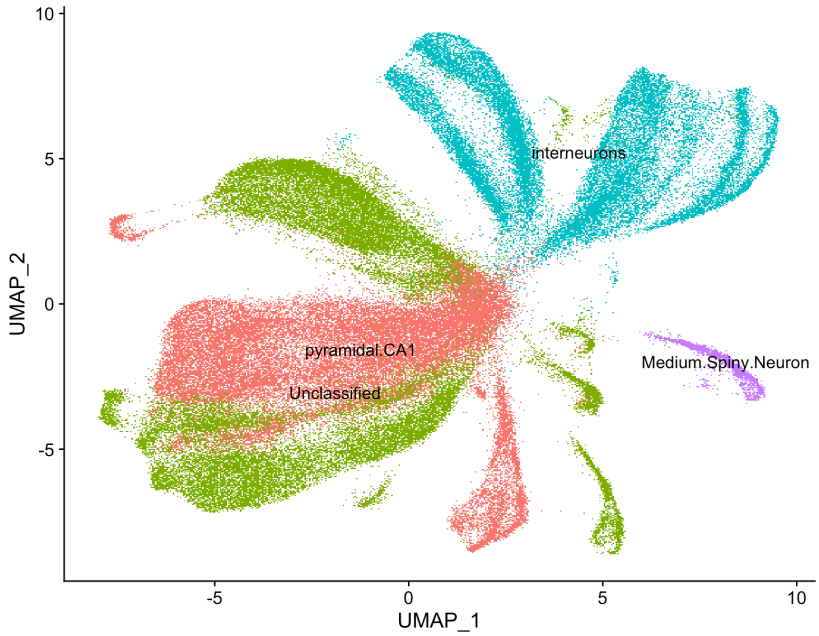
To explore if genes differentially expressed in neurons in PD, PDD and DLB arise from a global transcriptional change across all neuronal sub-types or from a specific subset of neurons, all neurons (inhibitory and excitatory) across all 4 conditions (PD, PDD, DLB, controls) were clustered using Louvain clustering implemented in Seurat. This identified 10 neuronal clusters (Figure 2.10a).

Among the 10 clusters, Cluster 2 consisted predominantly of nuclei from PD and control cases (specifically, 90% of nuclei from Control and PD subjects), suggesting an absence of nuclei from subjects with cortical Lewy body pathology (PDD and DLB) in this cluster (Figure 2.11 and 2.10a). In contrast, Cluster 6 consisted entirely of nuclei derived from non-DLB subjects (i.e., controls, PD, and PDD), suggesting these neurons are preferentially lost in DLB. To investigate this further, neuronal clusters were assigned to cell-type subtypes using a classification scheme based on Skene et al., 2018 [284] (Figure 2.10). Nuclei in cluster 6 were identified as coming from medium spiny neurons while the remaining clusters were identified as either pyramidal neurons, interneurons or unclassified. This suggests that medium spiny neurons may represent a vulnerable cell-type in DLB, and indeed unrelated research has suggested a selective dendritic degeneration of medium spiny neurons in the caudate nucleus in DLB subjects [354].

The reason for the apparent selective vulnerability of medium spiny neurons in DLB is unclear. Several genes identified as marker genes for medium spiny neuron (cluster 6), namely Cyclin-Dependent Dinase-Like 5 (CDKL5), Mitogen-Activated Protein Kinase 1 (MAPK1), Forkhead Box Protein P2 (FOXP2), Protein Phosphatase 2 Regulatory Subunit Beta (PPP2R2B), Protein Phosphatase 1 Regulatory Inhibitor subunit 1B (PPP1R1B), Insulinlike Growth Factor1 Receptor (IGF1R) and SCN1A are ubiquitously expressed across all neuronal types. In con-



(a) **A plot of all neuronal (excitatory and inhibitory) derived from all individuals.**



(b) **Cell-type identification of each cluster.**

Figure 2.10: Cell-type specific clustering of neurons.

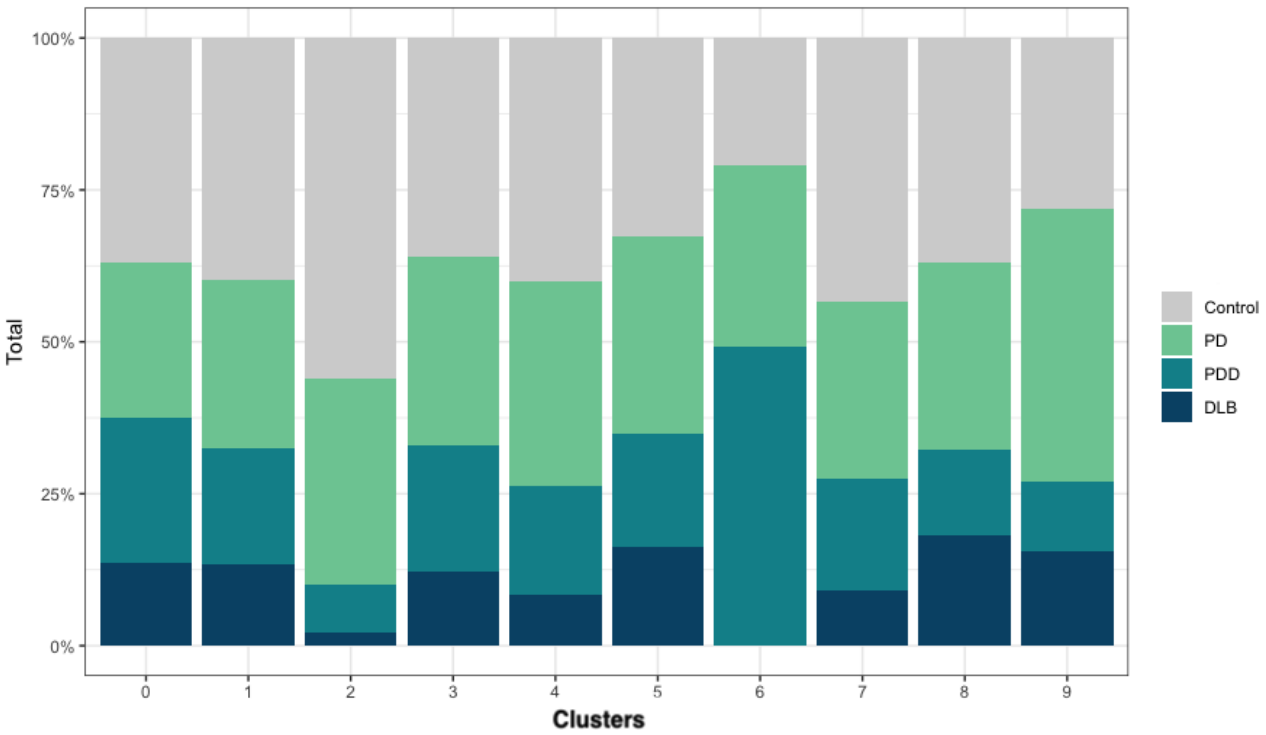


Figure 2.11: **Proportions of neurons derived from each group for each cluster.**

trast, PPP1R1B was highly specific to cluster 6 (medium spiny neurons), and the protein product of PPP1R1B is an inhibitor of PP-1 (Protein Phosphate-1), which is known to play a role in synaptic function and dopaminergic neurotransmission [184, 195, 228].

2.4 Discussion

This study applied snRNA-Seq to profile the gene expression of 21 Lewy Bodies disease and seven neurologically "healthy" cortical samples. Differential expression analysis was performed using a scRNA-Seq-specific method, MAST[89], which produced a set of DEGs across multiple cell-types in case-control and case-case-based comparisons. Overrepresentation analysis of the DEGs identified several pathways with a previously implicated role in the development of PD, including pathways related to metabolism, the immune system and signal transduction. Overall, transcriptional similarities and differences between LBDs and control were observed for multiple cell-types. More detrimental effects were observed in excitatory and oligodendrocytes in the dementia subjects (PDD and DLB) relative to PD.

2.4.1 Transcriptional altercations in multiple cell-types distinguishes the Lewy Body diseases

The significant points that emerged from the DE analyses were: (i) cell-type specific gene expression changes were identified across different comparisons; (ii) most of the DEGs were commonly dysregulated across multiple comparisons (including Lewy Body Disease (LBD)-control and LBD-LBD comparisons); (iii) over a third (36%) of DEGs were differentially expressed in cell-type disease-specific manner; (iv) several well studied and characterised PD-associated genes were also found to be dysregulated in dementia groups (PDD and DLB). Further, most were among the commonly dysregulated genes in the LBDs. Of the 21 genetically linked PD-genes considered , 13 were dysregulated in one or more comparisons (Figure 2.2). Although this was not surprising given the neuropathological and clinical overlaps between the LBDs, the direction of effect of some of these genes was not the same in PDD and DLB as observed in PD. For example, GBA, a gene that encodes a lysosomal enzyme, was only dysregulated in PDD (in excitatory neurons), even though previous genetic studies have identified mutations in GBA as common genetic risk factors in PD and DLB [74]. However, it has also been reported that GBA mutation carriers in PD cases have a reduced survival rate and an increased risk for early de-

mentia than noncarriers[321]. Further, a mutation of the glucocerebrosidase enzyme (encoded by GBA) was considered to induce the transition of the monomeric form of α-synuclein to an oligomeric, aggregated toxic form observed in PD [206, 297]. Similar results were observed for SNCA (up in PD, down in PDD and DLB), VPS35 (Vacuolar Protein Sorting-Associated Protein 35; up in PD, down in PDD and DLB), UCHL1 (up in PD and DLB, down in PDD), and DNAJC6 (down in PD and DLB, up in PDD). In contrast, examples of genes with the same direction of effects in the LBDs include Parkinsonism associated deglycase (PARK7), PINK1, SYNJ1 and Vacuolar Protein Sorting 13 Homolog A (VPS13A). Taken together, these observations suggest that, while it may appear at first that the two dementias are a continuation of PD, upon closer inspection, all three LBDs represent genetically heterogenous conditions with a distinct gene expression pattern of genetically linked synaptic or neuronal genes.

2.4.2 Selective vulnerability of medium spiny neurons in Dementia with Lewy bodies

The observation that excitatory neurons were more strongly impacted than other cell-types in the LBDs suggests a selective vulnerability of certain cell populations, but this effect could be confounded by the higher proportion of excitatory neurons in the samples. To address this issue, sub-clustering analysis was applied to dissect the cellular heterogeneity and identify different subpopulations of vulnerable neurons in the LBDs. This approach improved the resolution of the analysis, revealing previously unrecognised patterns of vulnerability. Furthermore, these "vulnerable" neurons were found to have a similar transcriptional profile to that of MSN, also known as GABAergic projection neurons. In a study, a selective loss of MSN was reported in the caudate nucleus regions in DLB cases [354]. Interestingly, in the current study, MSN-like neurons were present in PD and PDD subjects; therefore, it appears that the selective vulnerability of these neuronal types is unique to DLB.

2.4.3 Pathways perturbed across multiple-cell types

Functional enrichment analysis of DEGs using gene ontology (GO) terms corresponding to biological processes revealed several pathways perturbed in multiple-cell types. PDD and DLB

showed a higher degree of commonality (Figure 2.5). Most GO terms were commonly enriched for both down- and up-regulated DEGs. However, it was evident that:

1. Most significantly (FDR < 0.05) enriched terms were mainly enriched for down-regulated genes in a given dementias-control comparisons, and terms related to synaptic vesicle transport and response to oxidative stress were uniquely enriched for down-regulated genes in the two dementias in comparison to control and PD.

2. Up-regulated genes, which were identified primarily in comparisons of PD with control, were mainly enriched for genes involved in pathways related to metabolic processes, mRNA splicing and autophagy.

3. There were no significant (FDR < 0.05) terms found enriched for DEGs in comparisons of DLB with PDD. In summary, GO-terms-based enrichment analyses highlighted the commonalities between DLB and PDD, while PD exhibit pathological characteristics distinct from the dementia groups.

Similarly, genetically associated PD-related pathways were (i) mainly enriched for down-regulated genes in the dementias (PDD and DLB); and terms related to cell programmed death were uniquely enriched for down-regulated genes found exclusively in glial cell-types (Figure 2.8); (ii) while, up-regulated genes in comparisons of PD with control were enriched for terms related to transmission across synapses, chromatin organisation and autophagy-lysosomal systems(Figure 2.7).

Focusing on overlapping impaired pathways in the LBDs, the most noticeable difference between PD and PDD-DLB was the down-regulation of genes involved in the autophagy–lysosome system in the two dementias and their up-regulation in PD (Figure 2.5 and 2.6). This observation was partially explained by Kong et al. (2014) [158] in which the authors demonstrated the elevated expression of genes encoding transport ATPases (for example, ATP13A2 observed in PD may be brought about by the surviving neurons of the substantia nigra pars compacta in PD overexpressing these genes as a result of increased export of exosome-associated α-Synuclein [158]. Although contradictory results have been reported in the frontal cortex of PD and DLB

cases, decreased protein levels have been observed in disease cases compared to controls [221]. Furthermore, the expression of several genes that directly encode for autophagy-lysosomal components in excitatory neurons from PDD and DLB as compared to PD and control was marked by a shift in the range of the top 25% highest expressing nuclei and were highly expressed in sub-neuronal cluster 2 which mainly contained nuclei derived from PD and control subjects (Figure 2.11), suggesting these types of neurons may have been lost earliest in the progression of PDD and DLB.

2.4.4 Heritability enrichment analysis reveals a genetic association between differentially expressed genes and genetic determinants of Alzheimer's disease.

In genetic association analyses carried out using DEG sets, surprisingly, down-regulated genes in the LBDs were not enriched for genetic determinants of PD age of onset or PD risk and AD risk, suggesting that the observations may have been consequences of the disease process and not the cause. In contrast, up-regulated genes (from PD-control comparisons) were enriched for genetic risk of PD age of onset in astrocytes and OPC. Although it has previously been reported in OPCs [39, 4], the involvement of astrocytes in PD age of onset needs to be further explored since this may be the first study to have highlighted such a connection (Figure 2.9). Furthermore, several genetic associations were identified using 10% of most cell-type-specific genes in each disease group. While there was a nominal (FDR < 0.1) genetic association between PD genetic determinants and genes highly expressed in excitatory neurons, there was a significant (FDR < 0.05) enrichment of heritability for genes highly expressed in microglia and AD risk. These findings mirror similar observations reported in previous literature in which the majority of risk loci discovered through GWAS were highly expressed in microglia, which then led to the hypothesis that microglia play a role in the progression of dementia [230, 240].

2.4.5 Limitations of the study

A few limitations may complicate the interpretation of some of the observations made in this study. These limitations can be divided into technological – related to the tools and methods used- and non-technical – related to biological biases. Starting with biological limitations, there was a clear difference in the proportions of sexes between the four groups, although this is a well-known issue in the literature [225], although this could also be due to the low number of samples. Furthermore, the use of post-mortem tissue samples prevents comparing the magnitude of the difference between molecular events associated with the early stages of the disease and those that arise later. Focusing on the technical factors, except for sub-cell-type clustering analysis, all other findings presented in this chapter were based on snRNA-Seq data, which has a few challenges limiting the interpretability of the results. For example, although single-cell analysis provides a better view of the at a single-cell resolution, the tool used in the differential expression analysis is known to have a high false positive rate. However, based on a recently published study that analysed several tools used in single-cell analysis, MAST performed better than most available tools [290, 333]. Therefore, though DEA carried out using MAST provides a better result than pseudobulk analysis (which produced a significantly low number of DEGs ranging from 1-10), interpreting the results obtained from MAST should be done cautiously.

Chapter 3

Assessing gene expression changes associated with the earliest detectable abnormal Tau species

Abstract

Understanding the early cellular changes in AD is crucial for the development of effective disease-modifying therapies. This study aimed to investigate gene expression changes associated with early detectable abnormal tau species at the single-cell level. Brain samples with neurofibrillary tangles (NFTs) and tau oligomers were analysed, revealing transcriptional similarities and differences between these two tau species. Differential expression analysis identified several well-known neurodegenerative disease-associated genes, highlighting their dysregulation in the tau oligomer-positive (with few or no NFTs) group. Functional enrichment analysis revealed dysregulated pathways involved in protein aggregation processes, protein modification, mRNA splicing, and interleukin 12 (IL-12) signalling pathways. Additionally, the study observed the activation of reactive astrocytes in early tau pathology, suggesting their involvement in tau aggregate formation. The findings emphasize the importance of characterising tau oligomers in AD brain tissue to differentiate early-stage AD from established AD and avoid misclassification. In summary, this study provides valuable insights into the early molecular events in the formation of abnormal tau species, facilitating the development of targeted therapeutic strategies.

3.1 Introduction

3.1.1 Amyloid and Tau pathology in Alzheimer's disease

The German Bavarian neuroanatomist Alois Alzheimer first describedAD in 1907 when he presented his long-term study of a case involving a woman who suffered from what he described as "a severe disease process of the cerebral cortex" [28]. Since the original description, the neuropathological features of AD have included: (i) the amyloid plaques (first described by Alzheimer as "miliary foci"), which are extraneuronal aggregates of the protein amyloid beta (Aβ); (ii) and neurofibrillary tangles (NFTs), which are composed of aggregated filamentous tau protein, a microtubule-associated protein. The amyloid plaques are formed due to the abnormal processing of Amyloid Precursor Protein (APP) by beta- and gamma-secretases [241]. Two main amyloid plaques are primarily observed in AD pathology- the dense core and the diffuse [70]. Diffuse plaques form in the neuropil and generally lack neuritic components, while cored plaques are associated with dystrophic neurites. While NFTs in the brain of AD mainly consist of hyperphosphorylated and aggregated tau proteins, the lesions can also include pretangle material and neurophil threads [70].

3.1.2 The role of tau oligomers in disease

Since the initial description of AD, the underlying causes and pathogenic mechanisms of AD have remained largely elusive. There are currently many proposed AD pathogeneses, including the tau pathology hypothesis [241, 70], oxidative stress [42] and the amyloid cascade hypothesis [70, 72]. The tau hyperphosphorylation hypothesis proposes that abnormal phosphorylation causes tau to dissociate from microtubules and fold into paired helical filaments, the constituents of NFTs, presumed to be neurotoxic. Further, studies have shown the pathological features of tau pathology in AD are quite distinct from that of other tau-related diseases; for example, aggregates incorporate tau molecules carrying both 3 and 4 units of the microtubule-binding domain; and elevated levels of phosphorylated tau are observed in cerebrospinal fluid (CSF) of AD patients and tau pathology has been shown to correlate more closely with cognitive decline than amyloid pathology [71, 13, 365, 345].

Physiologically tau is an unfolded protein made up of four domains: (i) an acidic N-terminal domain (ii) a proline-rich region, (iii) an MT-binding, microtubule binding (MTBR), and (iv) a carboxy-terminal tail [37]. Functional tau is involved in actin reorganisation, microtubule stabilisation and neurite growth regulation [17]. Post-translational modifications, including ubiquitination, acetylation and phosphorylation are needed to regulate the physiological functions of tau proteins[15]. However tauopathies (a group of tau-related pathologies) are associated with excessive or abnormal phosphorylation [15], although it is not completely clear if phosphorylation is the cause or consequence of tau misfolding and aggregation, as tau multimerisation can be detected at early stages, even before several phosphorylated epitopes. In some animal studies, over-expression of tau caused cell death, synaptic dysfunction and behavioural abnormalities [351, 130]. Interestingly, Santacruz et al. (2005) [269] showed that despite the continuous formation of NFTs, a reduced tau expression in mutant transgenic mice decreases neuronal cell loss, suggesting NFTs formation may not cause neuronal loss.

Additionally, some studies have suggested that NFTs may not be the toxic species but instead the tau oligomers, an intermediate species made of soluble tau aggregates, may mediate neuronal toxicity [227]. These oligomers are formed when an individual tau binds to another, forming a soluble aggregate. In some studies, the levels of tau oligomers correlated with the clinical symptoms of AD suggesting that tau oligomers may be involved in the progression of AD [227]. Furthermore, in one study [99] tau oligomers derived from traumatic brain injury were shown to cause cognitive impairment, suggesting that oligomeric forms of tau may be involved in initiating the onset and spread of neurodegenerative diseases. Additionally, unlike tau monomers, tau oligomers were shown to inhibit axonal transport [147]. By far, the most noteworthy finding was the prion-like transmission mechanism by which tau misfolded conformers propagate in tau-related neurological disorders, in which tau oligomers may play a central role [15].

3.1.3 Tau seeding and propagation

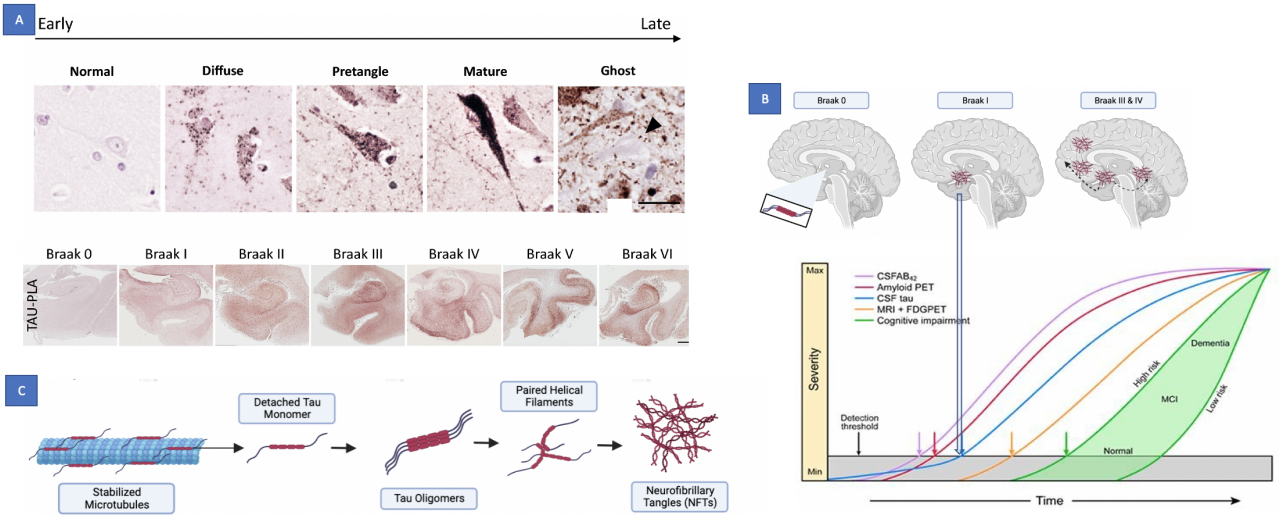


Figure 3.1: .
Insights into Alzheimer's Disease progression through visualisation of tau protein aggregates. **a** Top: Different staining patterns for tau using Tau-proximity ligation assay (tauPLA) from left to right: Normal refers to non-pathological state; Diffuse is characterised by the accumulation of small, diffuse tau protein aggregates; Pretangle refers to the early accumulation of tau aggregates; Mature stage is characterised by the formation of mature tangles that are densely packed, paired helical filaments of tau; Ghost pathology refers to the presence of tau protein aggregates that persist after the death of neurons. Scale bar 50 µm. Bottom: distribution of tau at different Braak stages. Scale bar 2 mm. **b** Alzheimer's disease progression. Blue arrow indicates detectable tau in cerebrospinal fluid (CSF) **c** Neurofibrillary tangle formation (Khanahmadi et al. (2015) and Bengoa- Vergniory et al. (2021))

The mechanism behind the prion-like spreading of tau has been studied using cell models that mimic tau seeding and propagation. *In vitro* studies have demonstrated the ability of tau to recruit soluble monomeric tau in a similar manner to prions [15]. As the disease progress, the accumulation of tau follows structurally linked pathways. Studies in animal models have revealed that endogenous tau misfolds and propagates from unaffected brain regions to affected brain regions through neuroanatomical networks. Usually from the entorhinal cortex and progress to the hippocampus, eventually to the surrounding cortex region (Figure 3.1) [3]. In addition, tau pathology propagates at different rates in different brain regions [149]. Collectively, these observations demonstrate the existence of abnormal tau species (different strains of tau) with distinct neuropathological and morphological characteristics [128]. Therefore, identifying and

characterising these tau "toxic" species is essential in designing therapies to lower the level of toxic tau, to decrease tau aggregation or to prevent the spread of pathology. Further, failure in identifying abnormal tau species, which do not carry the phosphorylated signature recognised by commonly used antibodies, could lead to a misclassification of diseased brain tissue (e.g. early stages of disease) as control, thus leading to false negative error.

3.1.4 Techniques for detecting tau species

Current methods used to identify misfolded tau on tissue sections are mostly dependent on the identification of phosphorylated tau epitopes with specific antibodies. Visualisation of phosphorylated tau has been used to develop staging schemes to classify the different phases of the pathology (Figure 3.1). Braak observed that pathological tau spreads throughout the brain in a stereotypical predictable pattern and categorised the progression of pathology in several Braak stages (I-VI)[33](Figure 3.1), currently the standard neuropathological staging of tau AD-related changes. In 2006, Braak et al. improved their staging method, originally based on the visualisation of NFTs with silver stain, by incorporating immunohistochemistry with AT8 antibody, which detects phosphorylated tau. Several other antibodies have been developed to visualise abnormal tau on tissue sections, recognising different phosphorylated tau epitopes or another biochemical characteristic of abnormal tau, such as the monoclonal Receptor Polyclonal Antibody (MC1) is an antibody, which recognise conformation-dependent epitopes in tau [140].

3.1.5 Using tau-proximity ligation assay to visualise early tau-tau interaction *in situ*

Despite the recent advancement in detecting phosphorylated tau species, a large portion of tau pathology is still not being detected using immunochemistry techniques. Recent studies have shown that tau-tau interactions occur early in the development of AD pathology (Figure 3.1), and that seed propagation of tau can happen in the absence of phosphorylation [150]. Therefore, visualising early tau-tau interaction could help identify early pathological tau-multimers and potentially representing early stages of AD. Bengoa-vergniory et al., 2021 [21] used a novel

tau-proximity ligation assay to directly visualise tau-tau interactions *in situ* regardless of tau phosphorylation state or conformational changes (Figure 3.1). The assay was able to recognise the multimers but not monomers. Furthermore, Bengoa-vergniory et al. were able to detect NFTs-like lesions and a small-sized diffuse pathology, which had not been reported before. The authors concluded that tau multimers were present from the earliest pre-symptomatic Braak stages [269, 21]. Most recent studies have looked at gene expression changes at late-stages of tau pathology including AD and were able to identify genes involved in the pathogenesis of AD. However, many late-stage clinical trials that were based on these studies were unsuccessful at treating AD, which has led to the recognition for the need to intervene at an early stage in the disease progression [84]. This was supported by studies showing that gene expression changes can occur even at the early stages of tau pathology [252, 154]. Therefore, it is believed that early intervention could potentially prevent the cascade of multiple pathologic pathways associated with tau pathologies and help in identifying early molecular changes that are critical for diagnosis and therapeutic strategies. Here we used snRNA-Seq analysis (to look at expression changes at the cell-type level, for example to assess the role of astrocytes in the disease progression) to identify gene expression changes occurring at the early stage of tau pathology.

3.1.6 Aims and objectives

The main aims of this study were:

1. To assess transcriptional signatures associated with oligomeric tau in the postmortem human brain using single-nucleus RNA-seq.

2. To evaluate and contrast transcriptomic similarities and dissimilarities between two distinct groups, brain tissues with tau oligomers (referred to as AT8-/tauPLA+) and those with tau oligomers and tau tangles (referred to as AT8+/tauPLA+).

3.2 Methods

3.2.1 Sample selection and tissue extraction

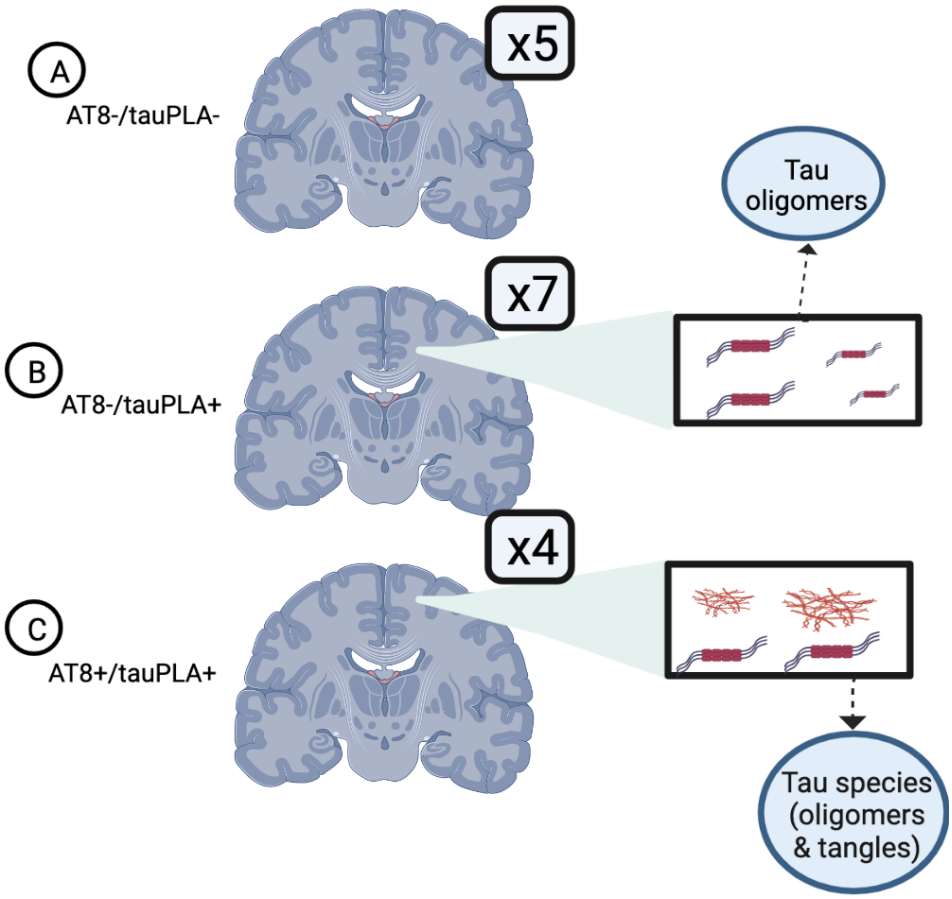


Figure 3.2: **Overview of samples used in the study.** 5 subjects were AT8 negative and tauPLA negative (A); 7 subjects were AT8 negative and tauPLA positive (B); 4 subjects were AT8 positive and tauPLA positive (C).

Brain tissue sample selection and preparation were performed by Dr Javier Alegre-Abarrategui and his group members using post-mortem samples donated to the Multiple Sclerosis and Parkinson's Brain Tissue Bank of Imperial College London. The samples were from temporal lobe neocortex, which sits halfway through in the Braak tau pathway between the early affected medial temporal lobe to the late affected primary sensorimotor cortices and can therefore harbour a range of tau pathology. Given that we aim at studying gene expression changes associated with the earliest tau species, and that these changes likely start at the pre-clinical phase of the disease process, all brain tissue samples were ascertained from donors with no

history of neurological disease antemortem (i.e., the pathological findings are in the absence of documented AD). Tau immunohistochemistry with AT8 antibody for phosphorylated tau and tauPLA was used to classify the temporal lobe samples of these cases into three groups (Figure 3.2). Immunohistochemistry for AT8 identified a group of samples with overt NFTs formation characterised by scattered NFTs. The remaining samples included cases with no tangles or rare isolated NFTs. This group was further subdivided with tauPLA into a group with negligible tauPLA signal (tauPLA negative) and a group with overt tauPLA diffuse signal (tauPLA positive) (Figure 3.2). The three sample groups were therefore:

- AT8 negative/tauPLA negative (double negative) – i.e., no abnormal tau deposition.
- AT8 negative/tauPLA positive (intermediate group) – i.e., tau oligomers but very little NFTs/hyperphosphorylated tau.
- AT8 positive/tauPLA positive (double positive) i.e., presence of both tau oligomers and NFTs / hyperphosphorylated tau.

To be consistent, a cut off of 0.78 of the quantification of tauPLA diffuse signal was used to define double negative groups (tauPLA- and AT8-). Clinical and pathological features of the cohort are available in Appendix B, Table 1.

Tissue extraction

Tissue sectioning was performed by Elisavet Velentza-Almpani under the supervision of Dr Javier Alegre-Abarrategui. Tissue sections of 5 µm thick Formalin-Fixed Paraffin-Embedded (FFPE) from the temporal cortex were used for tauPLA, and immunohistochemistry, and adjacent frozen tissue samples were homogenised in iced-cold PBS (10% w/v) using a Tissue Ruptor II with disposable probes (Qiagen) for snRNA-Seq analysis.

3.2.2 Immunohistochemistry

FFPE 5 µm thick human brain tissue samples from each individual were heated for 45 minutes at 70°C followed by dewaxing in xylene, rehydrated in 70-100% graded alcohols and finally blocked

with hydrogen peroxide (H_2O_2). Microwave heating was used to retrieve antigens (with citrate buffer pH 6 for 10 minutes). Samples were then incubated (after blocking the non-specific binding sites) with antibodies AT8, AT180, MC1 for 1 hour at room temperature. Antibodies were obtained from ThermoFisher Scientific (1:500 and 1:1000) and the MC1 antibodies were provided by Peter Davies, The Feinstein Institute for Medical Research, Manhasset, New York, United States. Next day, samples were incubated with biotinylated goat anti-mouse/anti-rabbit IgG secondary antibody for 1 hour at room temperature, then washed and incubated with Vectastain ABC reagents. Finally, slides were counterstained with Mayer's haematoxylin solution (Sigma-Aldrich), dehydrated with alcohols and xylene, and mounted with Distyrene, a Plasticizer, and Xylene (DPX) mounting reagent.

3.2.3 *In situ* tau proximity ligation assay

Following manufacturer's instructions, Duolink PLA kits (Sigma) was used to perform tauPLA. Tau5 antibody (ab80579, Abcam), along with Duolink PLA +/- kits, was used to prepare conjugated tau. Dewaxing was performed as mentioned in the previous section. Samples were incubated at 37 °C for 1 hour and diluted with tau conjugates at 4°C overnight. After being washed, the samples were incubated again with ligation stock and ligased for 1 hour at 37°C, washed again, followed by incubation with the amplification stock and polymerase for further 2.5 hours at 37 °C. As a final step, samples were incubated and washed again using detection solution for 1 hour at room temperature, washing and incubation step were repeated using substrate reagent for further 20 minutes at room temperature, counterstained for 5 minutes, dehydrated and finally mounted using DPX mounting reagent.

3.2.4 Neuropathological analysis

Aperio-Scanscope (40x objective) was used for imaging analysis. Three random samples were blindly taken as representative images. Neuropathological analysis was carried out using ImageJ software. For large perikaryal lesions (which consists mostly NFTs and neuropil threads (NTs)), a threshold of 12.5–100 μm^2 was used, while for small structures a threshold 1.5–3.5 μm^2 was

used. The average μm^2 was calculated for each image and semi-quantitative analysis was carried out by randomly sampling three representative images for each sample. Statistical analysis and comparison analysis was carried out using GraphPad Software. Finally, unpaired two-tailed Student's t test and one-way Analysis of Variance (ANOVA) test was carried out (statistical significance at $P < 0.05$).

3.2.5 Nuclei Isolation and sequencing

Isolation of nuclei for snRNA-Seq and RNA isolation was performed by Elisavet Velentza-Almpani. Library preparation and sequencing of snRNA-Seq data was performed by Source BioScience. A minimum of 5,000 nuclei per sample were targeted, and nuclei were sequenced at a depth of 30,000 paired-end reads per nucleus. All samples were sequenced on a single flow-cell to avoid sequencing batch effects using NovaSeq 6000. The libraries were prepared following 10X Genomic Chromium scRNA protocol. All samples passed all internal QC steps. Raw base calls were converted to FASTQ files (Phred+33 Illumina 1.9), a text-based format used to store biological data along with their quality scores and identification numbers. Adapter sequences, which are sequences ligated to the ends of cDNA fragments that can be introduced at the library preparation, were not detected. Therefore, adapter trimming step was not needed.

3.2.6 Analysis of single-nucleus data

3.2.7 Quality control

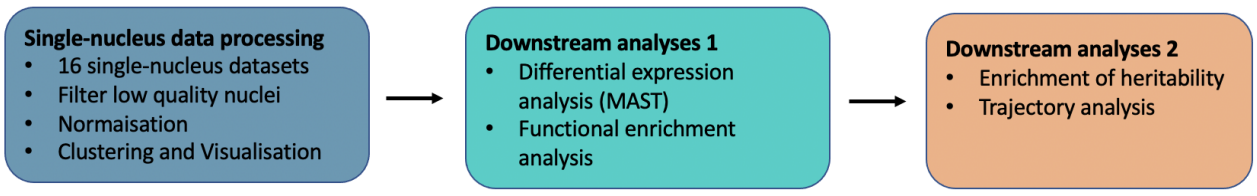


Figure 3.3: **Overview of pre-processing steps and downstream analysis carried out using 16 human brain tissue samples.**

FASTQ files were processed using `Cell Ranger`(v 3.1.0). Reads were mapped to both intronic and exonic regions using the human reference genome GRCh38. Genes were annotated using

Ensembl v93 and the final cell matrices (both raw and filtered) were generated.

Ambient RNA containing droplets (also known as empty droplets) were distinguished from droplets containing real nuclei using `CellBender` algorithm [91]. The following parameters were applied: (i) false positive rate (FPR) value set to 0.01; (ii) expected-cells (expected number of nuclei from each sample) set to 10,000; (iii) total-droplets-included (the total number of droplets from the rank-ordered UMI plot) set to 15,000. The final output, a new .h5 count matrix with ambient RNA removed, was used directly in the downstream analyses.

The next few QC steps were carried out using the R package `Seurat` pipeline [41]. Briefly, for each sample a Seurat object was created using the function `CreateSeuratObject()`. Genes detected in a minimum of 5 nuclei and nuclei with less than 10 % mitochondrial reads and at least 200 detected genes were retained. The R package `DoubletFinder` removed potential doublets from each sample. `DoubletFinder` [209] is a software that predicts doublets (technical artefacts or two or more nuclei that appear to be one) in snRNA-Seq data.

Default parameters were used: (i) PCs (statistically-significant principal components) value set to 10; (ii) pN (number of generated artificial doublets, expressed as a proportion of the merged real-artificial data) value set to 0.25; (iii) pK (a parameter used to compute the proportion of artificial nearest neighbours) value set to 0.09; (iv) nExp (total number of predicted doublets estimated from cell densities loads used in the 10x device) value varied per sample. `DoubletFinder` first simulates artificial doublets by calculating the average transcriptional profile of randomly selected nuclei pairs. Doublets are then identified by comparing artificial doublets profiles to "singlets" nuclei. Nuclei with a high proportion of artificial neighbours, calculated by dividing the number of artificial neighbours by the neighbourhood size, in gene expression space are removed from the dataset.

3.2.8 Integration of samples and cell-type identification

After removing doublets, an integrated data matrix was generated for downstream analysis. All individual samples were integrated using `Seurat` integration pipeline. First, each dataset was normalised for sequencing depth using `SCTransform` [112], a variance stabilising and trans-

formation tool from `Seurat`. `SCTransform` removes technical variations, including sequencing depth (the total UMI counts per cell), by modelling the counts using a regularised negative binomial model. The variance is adjusted by merging information across genes with similar abundances after the transformation. The primary purpose for variance adjustment is due to the fact that in scRNA-Seq data, cells with low UMI counts show disproportionately higher variance for highly abundant genes, reducing the variance contribution from other lowly abundant genes. Briefly, prior to integration `SCTransform()` function with the default parameters was used to normalise each sample. `SelectIntegrationFeatures()` function was used to select 3,000 features for integration. `PrepSCTIntegration()` function, followed by `FindIntegrationAnchors()` function, were used to identify anchors (cell pairwise correspondences between single cells). Finally, `IntegrateData()` function was used to generate an integrated assay dataset containing: (i) all UMI counts and (ii) centred and corrected Pearson residuals. The integrated assay dataset was visualised using UMAP after integration. Seurat functions `FindNeighbors()` and `FindClusters()` were used to compute the nearest neighbours and identify clusters using a shared nearest neighbour modularity optimisation based on the Louvain clustering algorithm. `FindAllMarkers()` function was then used to identify marker genes for each cluster (differentially expressed genes in each cluster compared to all other clusters). Each cluster was assigned a cell type using Fisher's exact test to examine the overlap between a cluster and a list of cell-type markers from Wang et al. (2008) [332].

3.2.9 Differential gene expression

In order to identify genes that were differentially expressed across the three groups, we applied MAST [89] for each pair-wise comparison (refer to section Chapter 1.1.2 for a detailed description of MAST). To explore gene expression differences at a single-cell level between the three groups (after adjusting for sex, age and individual ID), the following comparisons were undertaken:

i) AT8 negative/tauPLA negative (double negative) versus AT8 negative/tauPLA positive (intermediate group).

ii) AT8 negative/tauPLA negative (double negative) versus AT8 positive/tauPLA positive (double positive).

iii) AT8 negative/PLA positive (intermediate group) versus AT8 positive/tauPLA positive (double positive). (Table 3.1).

Analysis	Number of Subjects
AT8-/tauPLA- (double negative) vs AT8-/tauPLA+ (intermediate)	12
AT8-/tauPLA- (double negative) vs AT8+/tauPLA+ (double positive)	9
AT8-/tauPLA+ (intermediate) vs AT8+/tauPLA+ (double positive)	11

Table 3.1: Differential expression analyses

3.2.10 Functional enrichment of differentially expressed genes

Once differentially expressed genes were identified for each pair-wise comparison in each cell-type, `WEB-based GEne SeT AnaLysis Toolkit (WebGestaltR)` (v 0.4.4)[182] was used to extract underlying biological themes from the identified gene sets. `WebGestalt` is a widely used gene set enrichment analysis software tool that has integrated functional categories derived from centrally curated databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) and GO, and from locally curated databases originated from experimental and computational analyses. The R package of **WebGestalt**, `WebGestaltR` [182] was implemented using the default values, including overlap of 10 for minimum and maximum 500 genes and FDR-correction for multiple testing was performed using BH method and significant pathways identified using a threshold of FDR < 0.05. Briefly, gene sets were divided into up- and down-regulated DEG sets across the main cell types, resulting in fourteen gene sets per pairwise comparison. Each set was tested for enrichment against gene sets annotated with biological processes GO terms.

3.2.11 H-MAGMA and LDSC

Two statistical methods, H-MAGMA [275, 67], and LDSC [40, 90] were used to prioritise DEG sets associated with neurological phenotypes based on 3 publicly available GWAS summary statistics (refer to Appendix B, Table 6 and 7 for a full list of GWAS used). The methods account in different ways for confounders, including linkage disequilibrium and gene sizes. The first method, H-MAGMA is an extension of MAGMA which assigns intronic and intergenic SNPs to their nearest genes by incorporating chromatin interaction profiles from human brain tissue. Using default parameters, H-MAGMA was run in three parts:

1. Using GWAS summary statistics and H-MAGMA generated annotation files for an adult brain, association statistics at the gene-level were obtained by combining individual p-values. (H-MAGMA uses a window size 35-kb upstream and 10-kb downstream surrounding each gene, and LD is accounted for using data files from phase 3 of the 1000 Genome Project European population.) This step is used to quantify the association of each gene to a phenotype, while estimating gene correlations to reflect LD between genes. The output file includes information such as gene ID, gene mapping region (CHR, START, STOP), number of SNPs mapped to the gene, sample size for a gene, test statistic (ZSTAT and P-value), and r-squared (r^2) and adjusted r-squared values which explain the proportion of variance in the phenotype explained by the SNPs in a specific gene.

2. Using gene-level statistics created in the previous step, each gene set in the list of DEGs was then analyzed using a linear regression framework. The output file contains: the total number of genes in the gene set for the analysis; the linear regression parameters, including: β - the regression coefficient of the gene set, β_{STD} - obtained by dividing β by the standard deviation of the gene set (which can be used to compare the effect size of the gene sets), SE - standard error; and corresponding p-value.

3. Finally, significant gene sets are identified after running Bonferroni correction for the total number of analysed sets (significance level set at FDR < 0.05).

The second method LDSC, is a technique widely used to estimate heritability (Chpater 1.1.2). As mentioned in the previous chapter, LDSC estimates heritability by considering that SNPs in higher LD with other SNPs have higher test statistics on average for a polygenic trait/disease because of more causal SNPs being tagged. In other words, LDSC quantifies the contributions of polygenic effects and estimates its bias by assessing the relationship between LD scores and test statistics of SNPs from GWASs. The analysis was carried out as follows (after converting GWAS summary stats to LDSC format using (munge_sumstats.py)):

1. Annotation files were generated for each gene set using default parameters. SNPs were then mapped to genes using dsSNP file NCBI Build 37 co-ordinates (build 147 and hg19).

2. LD scores were computed for each annotation file using a 1 cM window (restricted to Hapmap3 SNPs), and the full baseline model was downloaded from the LDSC GitHub page (https://github.com/bulik/ldsc).

3. An enrichment score and its corresponding P-value were calculated based on the proportion of total SNPs per annotation, after taking into account all other annotations. Annotation categories with significant positive enrichment of SNP-heritability (tested using a one-tailed test) are then reported as a final result.

3.2.12 Single-nucleus trajectory analysis

Cellular asynchrony, in which a population of cells/nuclei captured are widely distributed in terms of progress, is a major problem that contributes to the apparent high variability of gene expression in snRNA-Seq data. To overcome this challenge, `Monocle` [46] was used to construct trajectories to identify changes in pseudo-temporal trajectories across the three groups. `Monocle` orders individual cells/nucleus in an unsupervised manner along a trajectory (for this analysis along tau oligomers load (A> B > C)) and assigns a pseudotime value to each nucleus that represents where the nucleus is along a given path. To identify genes driving differential trajectories and assess their biological implications, the following steps were taken:

1. Each cell type was analysed separately in order to extract cell-type specific driver genes.

2. Nuclei are clustered using Leiden community detection method with the default parameters implemented in `Monocle 3` function `cluster_cells()`. The function also calculates partitions (different branches), which are computed using a kNN pruning method and represent Leiden communities' superclusters.

3. Using `learn_graph()` function, the algorithm then learns the trajectory graph. This is done under the idea that in a high-dimensional space, a cell can be viewed as a point in space, and each dimension describes the expression of a gene. Therefore, by learning the trajectory the cells follow through that space, we can identify the program of gene expression changes. Since in snRNA-Seq data, there are many dimensions, `Monocle 3` uses the UMAP algorithm to reduce these dimensions.

4. The algorithm then uses `get_earliest_principal_node()` function to select a root node. Briefly, nuclei are grouped based on which trajectory they are nearest to. Then at each node, the fraction of nuclei coming from the earliest point is calculated. The node with the highest number of early nuclei (for this analysis, early nuclei are those belonging to group A) is returned as the root.

5. Once the root is defined, `Monocle 3` uses principal graph node IDs to specify the start of the trajectory. Further, using `order_cells()` function, a pseudotime value was assigned to each nucleus using the principal graph learned in the previous step and the position of the root state.

6. Next, genes that are differentially expressed across a trajectory were identified using spatial autocorrelation analysis (executed using the function `graph_test()`). Using Moran I test, nuclei with close proximity on a trajectory were tested to determine if they have similar gene expression for a given gene. The output of Moran I's test statistics include values from -1 to 1, with 0 representing no correlation, -1 anti-correlated and 1 perfect correlation. Significant genes (q-value < 0.05 and positive Moran I test statistics) were then selected for further analysis.

7. Using `find_gene_modules()` gene with similar expression (co-expressed) are then clustered into modules.

8. Finally, in order to gain more biological insight, functional enrichment analysis was run for each module identified in the previous step.

3.2.13 Code and Data availability

The raw and processed rat sequencing data generated in this study have not yet been deposited. All scripts used in this study can be found in GitHub urlhttps://github.com/rahfel. Supplementary materials including full tables can also be found in GitHub `https://github.com/rahfel`.

3.3 Results

3.3.1 Transcriptomic profiling of the temporal cortex

snRNA-Seq analysis was performed on 16 subjects with no history of brain disease. Each sample was allocated to one of three groups (Figure 3.2): 5 subjects with AT8-/tauPLA- (double negative), 7 subjects with AT8-/tauPLA+ (intermediate) and 4 subjects with AT8+/tauPLA+ (double positive). A diffusion score (quantification of tau-PLA labelled diffuse signal) of < 0.78 was used as a cut-off to define a **tauPLA negative** sample. Samples were matched for demographic factors; however, there was a difference in the proportions of sexes between groups due to sample unavailability (refer to Appendix B, Table 1 for sample demographics including age, sex and tauPLA diffuse score). The ratio of males in each group was as follows (male/female): 4/1 in double-negative groups, 2/5 in intermediate groups and 0/4 in double-positive groups. Overall, quality measures for read mapping and sequencing were high (Appendix B, Table 2). On average 87% of reads were mapped to the human genome version Genome Reference Consortium Human Build 38 (GRCh38-hg38). In total, snRNA-Seq was generated on 52,008 nuclei. Across the sample set, 17,483 genes were expressed (defined as genes detected ($0 >$ read) in at least 5 nuclei), with an average of 1, 923 genes detected per nucleus (Appendix B, Table 2 for full metrics).

3.3.2 Cell-type identification

The snRNA-Seq datasets were integrated (using `Seurat`) and an integrated plot (using UMAP) was generated to visualise clusters representing different cell-types (Figure 3.4). Clusters were assigned a cell type using a Fisher's exact test (FET) and a list of cell-type markers from Wang et al., 2018 [332]. In total, there were 18,492 nuclei identified as excitatory neurons; 4,823 as astrocytes; 3,670 as inhibitory neurons; 18,267 as oligodendrocytes; 2,535 as OPC; 3,671 as microglia and 550 as vascular (refer to Appendix B, Table 3 for a list of nuclei derived per subject).

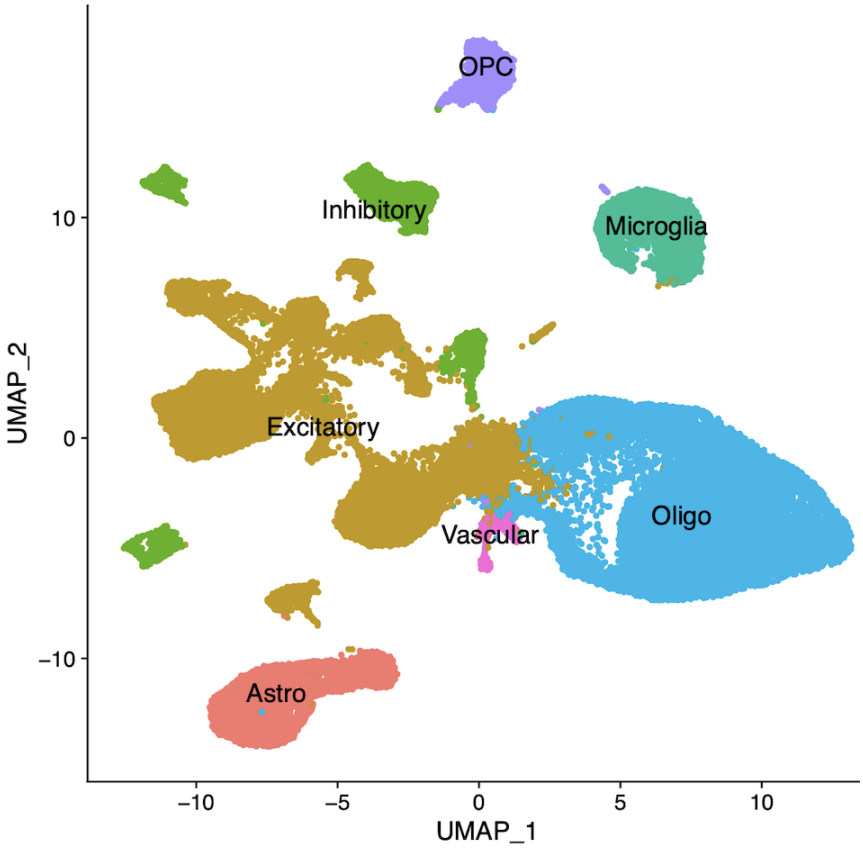


Figure 3.4: **A UMAP plot representing nuceli dervied from all subjects.** A joint plot was generated by combining approximately 52,000 nuclei. Abbreviations: UMAP, Uniform Manifold Approximation and Projection; Oligo, Oligodendrocytes.

3.3.3 Similar differential gene expression pattern was observed between intermediate and double positive groups

To explore cell-type specific transcriptional changes across the three groups, DEA was carried out in a pairwise manner across 6 broad cell-types (Table 3.1). There were significant (FDR $<$ 0.05) DEGs in all three pairwise comparisons across all cell-types. The largest number of DEGs were identified in AT8-/tauPLA- (double negative) versus AT8+/tauPLA+ (double positive) comparison. While, the least number of DEGs were identified in AT8-/tauPLA+ (intermediate) versus AT8+/tauPLA+ (double positive) comparisons (Figure 3.5; see Appendix B, Table 4 for a full list of genes DE in each comparison).

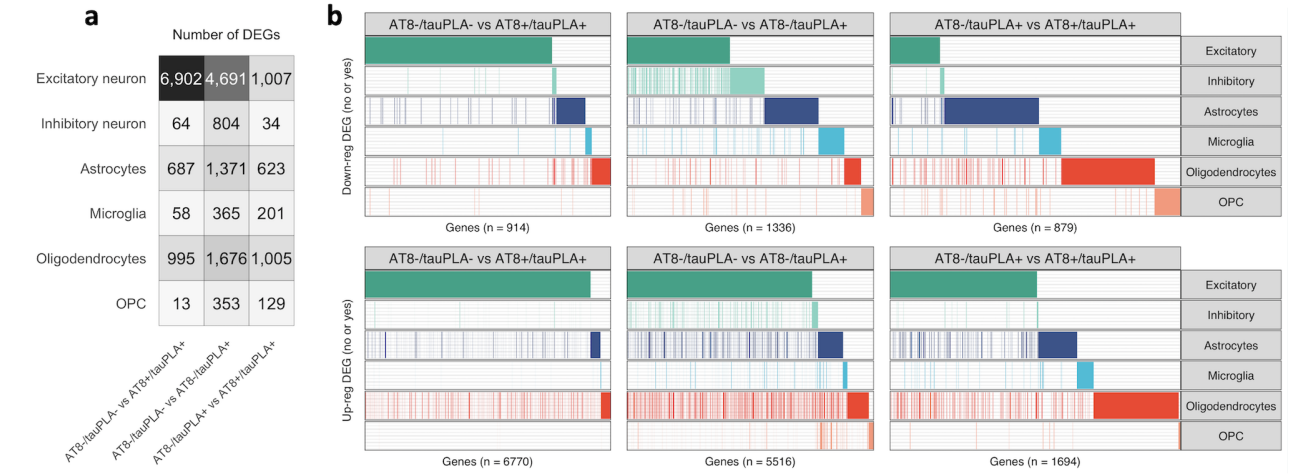


Figure 3.5: **Cell-type-specific gene expression changes and pathway across all comparison.** **a** Total number of differentially expressed genes (DEGs) across all pairwise comparisons (FDR < 0.05, $|\log_2(\text{fold change})| > \log_2(1.5)$). **b** A yes/no plot indicating if a gene (column) is expressed in a given cell-type comparison. The upper panel shows down-regulated genes, while the bottom panel shows up-regulated genes. The total number of DEGs (unique) in each comparison is indicated on the x-axis.

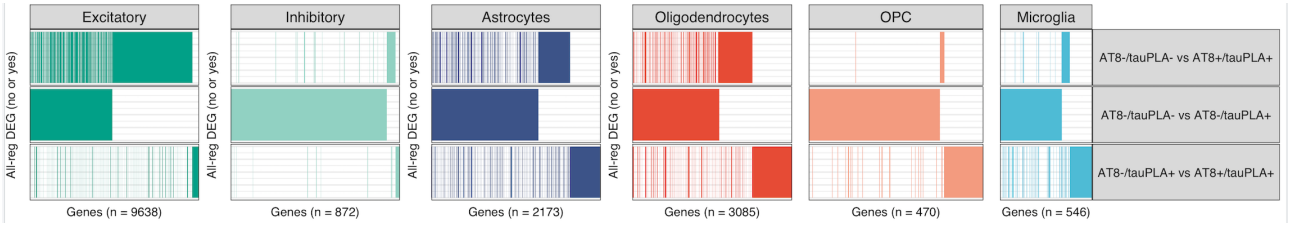


Figure 3.6: **Total number of differentially expressed genes across all cell-types.** (FDR < 0.05, $|\log_2(\text{fold change})| > \log_2(1.5)$). A yes/no plot indicating if a gene (column) is expressed in a given cell-type comparison. The total number of DEGs (unique) in each cell-type is indicated on the x-axis.

From the DEA carried out between nuclei isolated from AT8-/tauPLA- (double negative) versus AT8+/tauPLA+ (double positive) subjects, excitatory neurons showed a strong signature of down-regulation (i.e., 6,207 of DEGs were up-regulated (had higher expression level in double positive groups in comparison to double negative groups) while only 695 were down-regulated). While, in astrocytes, 552 genes were up-regulated while 135 were down-regulated. A similar trend was observed for AT8-/tauPLA- (double negative) versus AT8-/tauPLA+ (intermediate) comparison: 4,134 genes were up-regulated while 557 genes were down-regulated in excitatory neurons, while in astrocytes, 1,037 were up-regulated and 334 were down-regulated. Moreover, only a few numbers of genes were found to be differentially expressed in astrocytes and excitatory neurons in AT8-/tauPLA+ (intermediate) versus AT8+/tauPLA+ (double positive)

comparisons (Figure 3.5).

Overlap between cell-type specific gene sets from each comparison was visualised on a plot generated by the R package `ggplot` (Figure 3.5 and 3.6). Focusing on excitatory neurons, 50 % of DEGs overlapped between AT8+/tauPLA+ (double negative) versus AT8-/tauPLA- (double negative); and AT8+/tauPLA+ (double negative) versus AT8-/tauPLA- (double negative)(2,335 of 4,691 DEGs and 2,335 of 6,902 DEGs respectively). This observation may suggest similar neuronal pathways might be perturbed in the two tau pathologies. Further, for astrocytes, the overlap was 41% and 35% for oligodendrocytes (Figure 3.6).

3.3.4 Pathway enrichment analysis

To associate DEG sets with downstream functional consequences, over-representation analysis was carried out using the R package `WebGestaltR` [182]. In total, 694 biological process GO (child) terms were enriched for DEG sets derived from AT8-/tauPLA- (double negative) versus AT8+/tauPLA+ (double positive) comparison, while 420 and 81 GO (child) terms were enriched for AT8-/tauPLA- (double negative) versus AT8-/tauPLA+ (intermediate) and AT8-/tauPLA+ (intermediate) versus AT8+/tauPLA+ (double positive), respectively (Figure 3.7; refer to Appendix B, Table 5 for a list of significant GO child and corresponding parent terms across all comparisons).

Focusing on terms that were uniquely enriched for the up-regulated genes (mainly in excitatory neurons) in AT8+/tauPLA+ (double positive) group in comparison to AT8-/tauPLA- (double negative) group were terms related to:

1. Mitochondrion organisation, including mitochondrial electron transport, NADH to ubiquinone (FDR = 3.28×10^{-09}), protein targeting to mitochondrion (FDR = 3.12×10^{-06}) and regulation of autophagy of mitochondrion (FDR = 3.14×10^{-05}).

2. Ubiquitination, including protein polyubiquitination (FDR = 3.05×10^{-08}), histone ubiquitination (FDR = 1.48×10^{-02}), and positive regulation of protein ubiquitination (FDR = 3.82×10^{-03}).

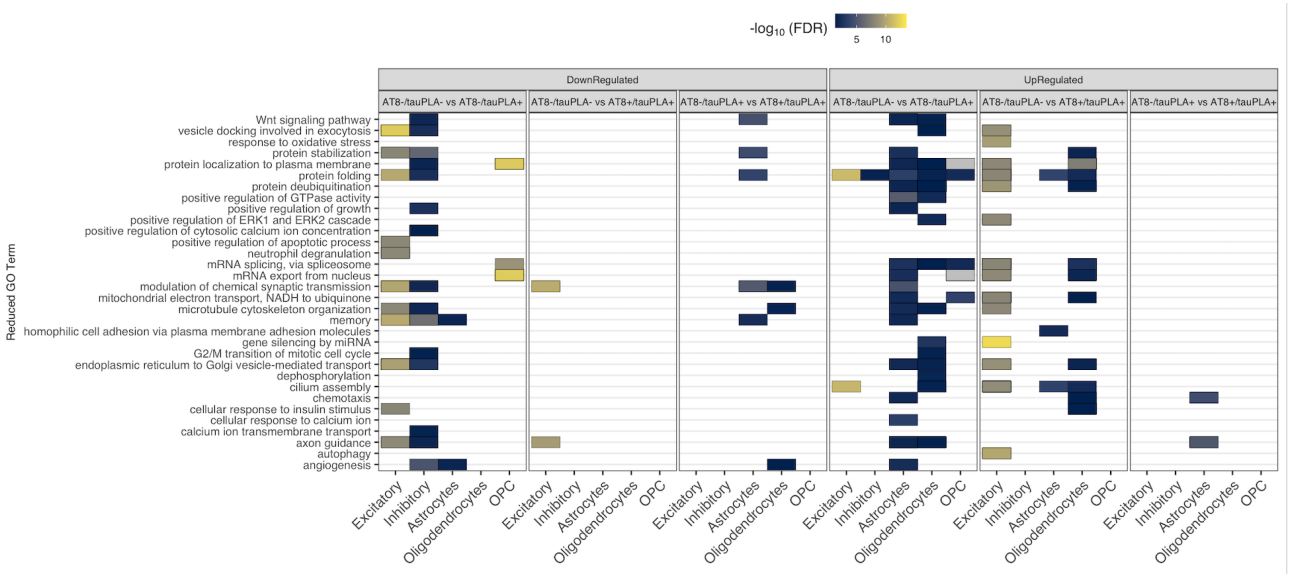


Figure 3.7: **Reduced biological process gene ontology (GO) terms** (referred to as parent terms) enriched for cell-type specific differentially expressed genes (up or down-regulated divided into two columns). The fill of each tile indicates the -log10FDR value of the most significant child term associated with the parent term. The full list of child GO terms assigned to each parent term across all pairwise comparisons in the panel is available in Appendix B, Table 5.

3. Interleukin-12 related pathways, including cellular response to interleukin-12 (FDR = 6.80×10^{-03}), interleukin-12-mediated signaling pathway (FDR = 1.65×10^{-02}), and response to interleukin-12 (FDR = 4.21×10^{-03}).

Among pathways enriched for down-regulated genes in the same comparisons were pathways related to cognition, including memory (FDR = 6.63^{11}), cognition (FDR = 4.92×10^{-12}) and regulation of system process (FDR = 2.34×10^{-04}).

3.3.5 Heritability enrichment in differentially expressed genes

Genetic enrichment of association (heritability enrichment) to neurological disease among the genes DE between the various comparisons was assessed using `HMAGMA` and `LDSC` (Figure 3.8). The aim was to test if cell-type specific DEGs from the three comparisons contribute to the common SNP heritability of AD, PD, and IQ. DEGs were split into up and down-regulated gene sets and tested separately (to provide information on the potential direction of effect). Using DEG sets from AT8-/tauPLA- (double negative) versus AT8+/tauPLA+ (double positive)

comparisons, the following were observed:

1. A significant (FDR < 0.05) association observed between AD genetic risk and genes up-regulated in oligodendrocytes ($\text{FDR}_{\text{LDSC}} = 4.09 \times 10^{-02}$, $\text{FDR}_{\text{HMAGMA}} = 0.25$) and excitatory neurons ($\text{FDR}_{\text{LDSC}} = 7.22 \times 10^{-09}$, $\text{FDR}_{\text{HMAGMA}} = 0.14$).

2. A significant (FDR < 0.05) association observed between PD genetic risk genes and up-regulated genes in excitatory neurons ($\text{FDR}_{\text{LDSC}} = 5.28 \times 10^{-06}$, $\text{FDR}_{\text{HMAGMA}} = 0.71$).

3. A significant (FDR < 0.05) association observed between IQ genetic risk genes and down-regulated genes in excitatory neurons ($\text{FDR}_{\text{LDSC}} = 1.29 \times 10^{-05}$, $\text{FDR}_{\text{HMAGMA}} = 0.09$).

Using DEG sets from AT8-/tauPLA+ (intermediate) vs AT8-/tauPLA- (double negative) comparisons, the following were observed:

1. A significant (FDR < 0.05) association observed between AD genetic risk and genes down-regulated in oligodendrocytes ($\text{FDR}_{\text{LDSC}} = 2.53 \times 10^{-03}$, $\text{FDR}_{\text{HMAGMA}} = 0.61$) and excitatory neurons ($\text{FDR}_{\text{LDSC}} = 3.47 \times 10^{-02}$, $\text{FDR}_{\text{HMAGMA}} = 0.24$).

2. A significant (FDR < 0.05) association observed between IQ genetic risk and genes down-regulated in astrocytes ($\text{FDR}_{\text{LDSC}} = 1.96 \times 10^{-04}$, $\text{FDR}_{\text{HMAGMA}} = 0.18$, oligodendrocytes ($\text{FDR}_{\text{LDSC}} = 4.94 \times 10^{-04}$, $\text{FDR}_{\text{HMAGMA}} = 0.18$) and excitatory neurons ($\text{FDR}_{\text{LDSC}} = 3.98 \times 10^{-02}$, $\text{FDR}_{\text{HMAGMA}} = 0.15$).

No significant association (FDR < 0.05) between any of the genetic risks and DEGs from AT8-/tauPLA+ (intermediate) versus AT8+/tauPLA+ (double positive) comparisons across all cell-types. A full list of analysis statistics, including `LDSC` and `H-MAGMA` based analyses can be found in Appendix B, Table 6 and 7.

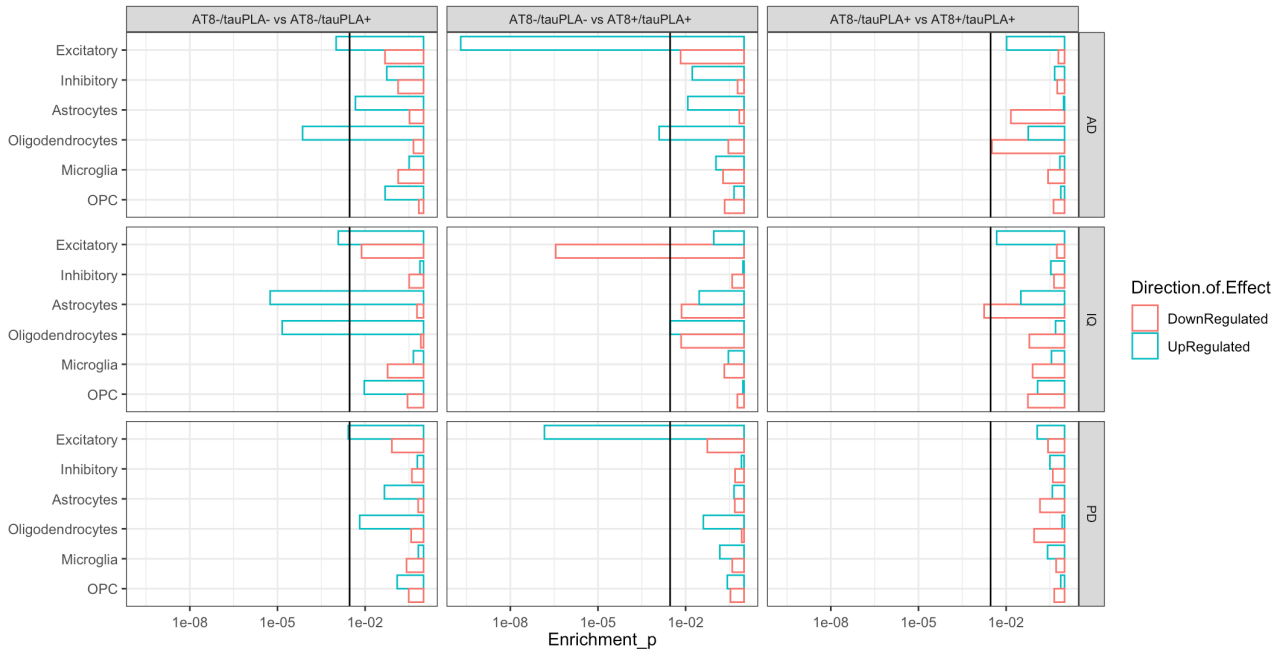


Figure 3.8: **Genetic association with cell-type specific differentially expressed genes across all pairwise comparisons**. LDSC was used to identify associations. The x-axis indicates enrichment p-values. The black line indicates Bonferroni significance threshold (p-values adjusted for the number of cell-types tested; FDR < 0.10). The colour of the bars indicates if the DEGs were up or down-regulated. Abbreviations: AD, Alzheimer's disease; PD, Parkinson's Disease; IQ, Intelligence quotient. Full results can be found in Appendix B, Table 6

.

3.3.6 Trajectory analysis

Using `Monocle 3` [46], graph-autocorrelation analysis was carried out to identify co-regulated sets of genes (also referred to as modules) that vary (i.e., have a trajectory) across the natural ordering of the groups from AT8-/tauPLA- (double negative) > AT8-/tauPLA+ (intermediate) > AT8+/tauPLA+ (double positive).

Focusing only on astrocytes, `Monocle 3` generated 18 modules, with each module consisting of several genes (ranging from 231 to 1,048; Figure 3.9a). To determine whether any of the modules contain genes that might play a role in the development of AD pathology, functional enrichment analysis (using over-representation analysis as implemented in the R package `WebGestaltR`) was carried out (Figure 3.9c). Among the modules, module 7,9, 11,12 and 18 had the largest number of enriched biological process GO (child) terms (50, 51, 51, 50 and 49, respectively; refer to Appendix B, Table 8 and 9 for genes per module and enrichment of pathways result).

In addition, a one-sided Fisher's exact test was used to determine if any of these modules are enriched for reactive astrocytes, a population of astrocytes previously reported to be involved in the pathogenesis of several neurodegenerative diseases [233, 180]. Among the modules, module 11 was found to be enriched (Fisher's P-value = 6.12×10^{-03}) for reactive astrocyte marker genes (p-values for other modules can be found in Appendix B, Table 10).

Biological process GO terms related to innate immune response (FDR = 5.67×10^{-10}), regulation of cytokine production (FDR = 1.60E-06), cytokine-mediated signalling pathway (FDR = 1.72×10^{-03}), defence response to virus (FDR = 8.84×10^{-05}) and ageing (FDR = 8.96×10^{-03}) were found to be enriched for module 11 genes (refer to Appendix B, Table 9 for a full list of significant child and parent terms).

Furthermore, enrichment of genetic association analysis revealed a significant genetic association (FDR < 0.05) between gene sets in modules 5, 6, 9 and 14 and AD genetic risk (FDR = 2.73×10^{-04}, FDR = 4.76×10^{-02}, FDR = 2.08×10^{-02} and FDR = 5.05×10^{-04}, respectively; Figure 3.9b). Functional enrichment analysis for these modules showed significant enrichment for biological process GO terms related mRNA splicing (module 5, FDR = 6.62×10^{-05}; module

a

Module Number	Number of Genes
1	1,048
2	1,026
3	918
4	913
5	890
6	850
7	850
8	839
9	785
10	782
11	731
12	662
13	635
14	632
15	526
16	417
17	377
18	231

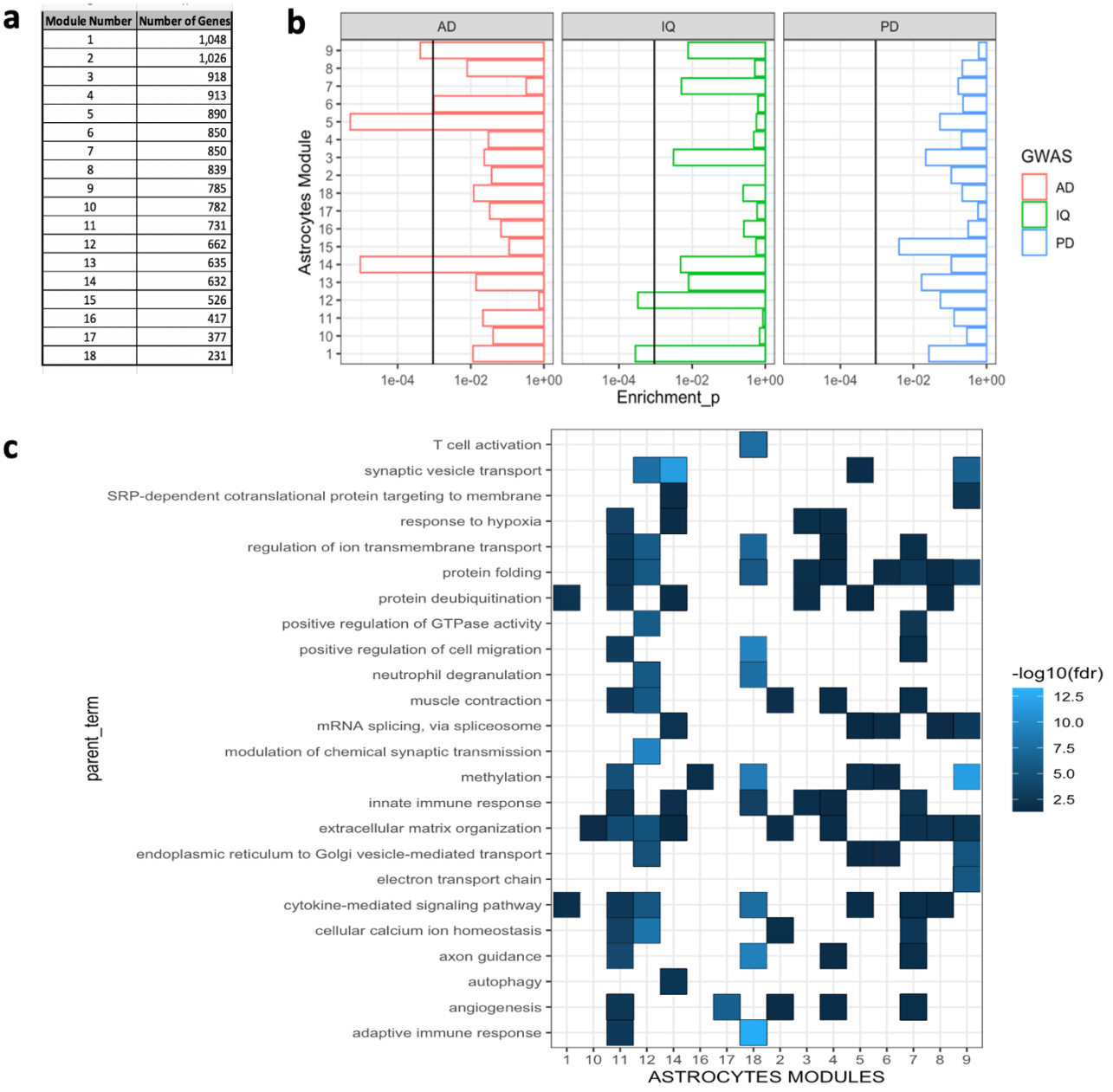


Figure 3.9: **Monocle 3 identified astrocytic gene sets referred to as modules**. **a** Number of genes per module. **b** Genetic associations carried out using LD score regression (LDSC). **c** Functional enrichment analysis using gene sets from each astrocytic module. Reduced (parent) Gene Ontology (GO) terms of the top 50 significantly enriched (FDR < 0.05) child terms per cell-type identified using overrepresentation analysis. Full overrepresentation analyses results and LDSC results can be found in Appendix B, Table 9 and Table 11, respectively.

9, FDR= 3.99×10^{-03}; module 14 FDR= 6.03×10^{-03}) and synaptic vesicle transport (module 5, FDR = 2.45×10^{-02}; module 9, FDR = 4.54×10^{-12}; module 14, FDR= 2.30×10^{-12}; Appendix B, Table 11).

Focusing on excitatory neurons, `Monocle 3` generated 20 modules containing between 312 to 1,415 genes. Genetic association test revealed a significant association between AD genetic risk and module 3 and 11 genes. Furthermore, overrepresentation analysis of module 3 and 11 genes revealed significant (FDR = 8.26×10^{-03} for module 3 and FDR = 2.22×10^{-02} for module 11) enrichment for terms relating to sodium ion transport (module 3, FDR = 3.00×10^{-02}), calcium ion homeostasis (module 3, FDR = 1.60×10^{-02}) and negative regulation of proteolysis (FDR = 2.95×10^{-03}). A full list of result tables for excitatory neurons can be found Appendix B, Table 12-14.

3.4 Discussion

Despite decades of research, the knowledge of early cellular changes in AD remains unknown. This knowledge gap has made it difficult to design effective disease-modifying or preventing therapies. Recent studies have shown that early tau oligomers are toxic species that may be responsible for the spread of tau pathology [227]. The current study used brain tissue samples with NFTs and brain tissue samples with tau oligomers to assess gene expression changes associated with the earliest detectable abnormal tau species at the single-cell level.

3.4.1 Differential expression analysis highlights transcriptional similarities and differences between neurofibrillary tangle and early detected tau multimers

Differential expression analysis revealed several well-known and characterised neurogenerative disease-associated genes (50/57) derived from Genomics England Gene Panel (Adult-onset neurodegenerative disorder V 4.13; refer to Appendix B, Figure 1 for overlap plot) were differentially expressed in AT8-/tauPLA+ group compared with AT8-/tauPLA- group. Example of these includes APP, LRRK2, PINK1, SNCA, MAPT, PSEN1, PSEN2 and PARK7. Interestingly, most of these genes were not differentially expressed in any comparison carried out between the two tauPLA positive groups (AT8+/tauPLA+ (double positive) in comparison to AT8-/tauPLA+ (intermediate)). Furthermore, fewer genes were found to be differentially expressed between the two tauPLA positive groups indicating transcriptional similarities. This observation mirrors the low number of significantly enriched terms identified using DEGs from the two tauPLA positive group comparisons (Figure 3.7).

3.4.2 Dysregulated neuronal and non-neuronal pathways identified in early tau pathology

Functional enrichment analysis revealed DEGs in the tauPLA positive groups (AT8-/tauPLA+ and AT8+/tauPLA+) compared with double negative (AT8-/tauPLA-) group were highly en-

riched for pathways involved in:

1. Protein aggregation processes including tau and amyloid β. Examples of these pathways include phosphorylation, splicing, autophagy of mitochondrion, memory and ubiquitination [186, 315, 232].

2. The process of modifying proteins by removing ubiquitin or ubiquitin-like molecules. Deubiquitination was among the pathways significant in up-regulated genes in tauPLA positive groups. Deubiquitinating enzymes (DUBs), which play a role in reversing the action of ligases by removing ubiquitin chains, are involved in several processes including synaptic plasticity, axonal growth, modulate mitochondrial clearance and neuronal cell fate [6, 185]. Dysregulation of certain DUBs has been found in several neurodegenerative diseases [6]. Consistent with the literature, some of these DUBs, mainly Ubiquitin Specific Protease (USPs), protease DUBs were dysregulated in the tauPLA+ groups, specifically in excitatory neurons and astrocytes. The role of USPs in glial cells is currently being investigated in various neurodegenerative disease models [139, 185]. USPs are involved in removing post-translational modification ubiquitin, thus preventing the degradation of several overexpressed proteins including tau [336]. For example, USP14 is involved in removing ubiquitin from substrate proteasome targets such as TAR DNA-binding protein 43 (TDP43), thereby rescuing the protein from degradation; and the inhibition of USP14 by 1-[1-(4-fluorophenyl)-2,5-dimethylpyrrol-3-yl]-2-pyrrolidin-1-ylethanone (IU1) was shown to enhance the degradation of proteins related to neurodegenerative diseases [172].

3. The process of mRNA splicing. Dysregulation of RNA Splicing in tauopathies has been reported in recent studies[125]. In the current study, genes involved in mRNA splicing were observed to be dysregulated in AT8-/tauPLA+ samples. This may indicate the role of tau oligomers in the disruption of the spliceosome.

4. Interleukin 12 (IL-12) -signalling pathways. Interestingly, a recent study has shown that Inhibition of IL-12 signalling reduced Alzheimer's disease-like pathology and cognitive decline [329]. In the current studies, genes that were up-regulated in tauPLA+ groups

compared with tauPLA- group were enriched in IL-12 signalling pathways.

3.4.3 Activation of reactive astrocytes observed in early tau pathology

Among the identified astrocytic modules, module 11 was enriched for reactive astrocytes marker genes. Reactive astrocytes (A1) are astrocytes that get activated as a response to abnormal events in the CNS [233, 180]. Unlike normal activated astrocytes as a result of being consistently activated by signals in the CNS, reactive astrocytes are activated due to pathological stimuli. In response to pathology (neurodegeneration, trauma, infection, etc), reactive astrocytes are reported to be involved in molecular programs, including transcriptional regulation, and physiological and morphological remodelling. Moreover, studies have highlighted the role of A1 astrocytes in removing dysfunctional neurons to preserve neuron circuit function and killing virally infected neurons to prevent the spread of viruses [233, 180]. In the current study, marker genes for A1 astrocytes are highly expressed in tauPLA+ groups (including AT8-/tauPLA+ group), suggesting that reactive astrocytes have an earlier role in the formation of tau aggregates than previously reported.

3.4.4 Designing effective AD transcriptional analyses/ practical lessons sample selections

Surprisingly, until now, no single macroscopic feature has been used to diagnose AD. However, what is now considered the standard for pathologic diagnosis is a combination of features which taken together, are highly suggestive of AD. These features include not only the extracellular Aβ and intracellular NFTs but also tau-positive neuropil threads and dystrophic neurites. Several histological stains and immunohistochemical methods have been developed to microscopically examine AD in multiple brain regions. However, most of these methods rely on the presence of hyperphosphorylated tau or neurofibrillary tangles, which are events appearing in the late stage of the disease. Therefore, this study shows how a novel *in vitro* assay, tauPLA which allows specific histological visualisation *in situ* of tau oligomers which are considered early molecular

events in AD. These features have a similar transcriptional profile as pre-AD samples with numerous NFTs and Aβ load. Further, (assuming tau oligomers as true toxic species) failure to characterise tau oligomers in samples could potentially lead to misclassification of AD brain tissue as control, therefore increasing the likelihood of false-negative or positive findings or a failure to differentiate early-stage AD from established AD.

Chapter 4

Integrative genomics reveals pathogenic mediator of valproate-induced neurodevelopmental outcomes

Abstract

Prenatal exposure to sodium valproate (VPA), an anti-seizure medication, is associated with an increased risk of adverse neurodevelopmental outcomes in children. This study aims to elucidate the molecular mechanisms underlying these consequences using integrative genomics. A rat model of VPA teratogenicity was utilised to mimic chronic oral VPA treatment during pregnancy at therapeutically relevant doses. Gene expression analysis was performed on the brains of VPA-exposed rat pups, revealing substantial differential gene expression and dysregulated splicing, independent of neuronal gain or loss. Pathway analysis demonstrated that VPA downregulated genes related to neurodevelopment and synaptic function, which were also significantly associated with heritability of human intelligence, schizophrenia, and bipolar disorder. These findings establish a mechanistic link between chronic prenatal VPA exposure and neurodevelopmental disability through transcriptional dysregulation.

4.1 Introduction

Neurodevelopmental disorders (NDDs) are a group of heterogeneous disorders that disturb the development of the central nervous system, which encompasses the brain and the spinal cord. It is estimated that >3% of children worldwide suffer from some form of NDDs [238]. As mentioned in the previous chapter, the prevalence of NDDs as defined by DSM-5 is 0.7-17% under 18 years old worldwide [92]. NDDs impair cognition, communication, behavioural, motor and daily living skills [238]. Studies have suggested that several factors contribute to the disease aetiology, including genetics variations, epigenetic mechanisms (which involve perturbations that disrupt the expression of risk genes) and environmental factors, or a combination of any of these factors [238, 268, 236]. Similar to neurodegenerative diseases, NDDs are characterised by multiple clinical features, and comorbidity is frequently reported in individual patients [236].

4.1.1 The genetic architecture of neurodevelopmental disorders

The genetic factors that determine and influence NDDs phenotypes consist of different mutations, including copy number variations (CNVs), small insertions or deletions (indels), point mutations and chromosomal rearrangements [341]. Identifying and characterising the potential genetic basis of NDDs is essential for understanding the functional and molecular mechanisms responsible for the onset and progress of these disorders. So far, several potential NDDs risk genes have been identified through GWAS, RNA-Seq-based studies and phenotype-genotype correlation studies. More importantly, family-based studies, including monozygotic twin studies, have been used to identify genetic risk factors as well as other environmental risks and protective factors that influence NDDs [341, 295].

Furthermore, GWAS, whole-exome sequencing studies and other functional studies have shown that common and risk variants associated with NDDs were found in genes involved in conserved functional molecular pathways, including chromatin remodelling, protein synthesis, epigenetic regulation, synaptic signalling and mRNA splicing [298]. Examples of NDDs-associated chromatin remodeler or splicing regulator genes, include chromodomain helicase DNA binding protein 1 (CHD1), chromodomain helicase DNA binding protein 8 (CHD8), DS cell adhesion

molecule (DSCAM), ASH1 like histone lysine methyltransferase (ASH1L), AT-rich interactive domain-containing protein 1B (ARIDIB) and Lysine [K]-specific Methyl Transferase 2A (KMT2A) [298]. Other examples of NDD-associated genes involved in mRNA splicing include SON DNA binding protein (SON) [298, 155], RNA Polymerase III Subunit B (POLR3B) [265], Transcription Factor 4 (TCF4), Transcription Factor 20 (TCF20) and Lysine demethylase 6B (KDM6B) [268, 271].

4.1.2 Drug-induced neurodevelopmental disorders

Over the last decades, cross-sectional and longitudinal human studies have shown that exposure to harmful stimuli in the womb during crucial postnatal developmental stages leads to short or long-lasting changes in brain structure. Recent studies using animal and human models of NDDs have shown harmful substances such as alcohol [170], caffeine [360, 23], nicotine, cocaine [340], amphetamine, ecstasy, opiates, and prescription drugs such as anti-psychotics and antiepileptic drugs [344, 53, 65, 335] produce alterations in neurodevelopmental trajectories.

Among the proposed underlying molecular mechanisms for drug-induced cognitive impairments, the role of the drug (methamphetamine) treatment on the disruption of cAMP signalling and glutamate regulation was identified in human embryonic stem cells (hESCs) derived from cerebral organoids [296]. Furthermore, RNA sequencing was used in a similar study to show a substantial down-regulation of genes involved in neurogenesis and neuronal system development in the drug-treated organoids compared with control organoids [64]. Another study used cortical organoids from human induced pluripotent stem cells (hiPSCs) to show transcriptional similarities between antiseizure drug-exposed organoids and autism patient-derived organoids, as well as postmortem ASD brain [216].

Antiepileptic drugs

Antiepileptic drugs (AEDs) are a diverse group of drugs used to prevent epileptic seizures; however, they are also commonly used to treat non-epileptic CNS disorders, including schizophrenia, migraine, neuropathic pain, anxiety, myotonia (described as a failure of muscle relaxation after activation) and bipolar affective disorder [215, 258]. Based on the specific molecular target in

the brain and the drugs' mechanism of action, AEDs can broadly be divided into four groups:

1. neuronal ion channel blocking AEDs, including phenytoin, carbamazepine, lamotrigine, and ethosuximide. Many of AEDs ion channel targets overlap with known human epilepsy genes, suggesting that genetic mutations and the AEDs have opposite effects on the function of these channels. For example, while gain-of-function mutations in the Calcium channel, voltage-dependent, T type, alpha 1H subunit (CACNA1H) could potentially increase seizure susceptibility by increasing the propensity of neurons to fire action potentials [337, 249, 51], the AED ethosuximide acts by inhibiting these channels (specifically T-type calcium channel) in the thalamus [258, 215].

2. AEDs which enhance inhibitory gamma-aminobutyric acid A (GABAA) receptor-mediated synaptic interactions. For example, benzodiazepines, such as diazepam and phenobarbital, are AEDs that facilitate binding inhibitory neurotransmitter GABA with its receptors, thus increasing the chloride channel opening frequency and enhancing synaptic inhibition [258].

3. AEDs that are glutamate α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor antagonists. While GABAA receptor-targeting drugs mediate synaptic inhibition, AEDs such as perampanel inhibit AMPA receptor-mediated synaptic excitation.

4. AEDs with mixed mechanisms including felbamate and sodium valproate [215].

Valproic Acid

Sodium valproate (VPA) is an AED widely used in the treatment of epilepsy, being the most effective drug to treat seizures in Genetic Generalized Epilepsy (GGE) [203, 202]. Furthermore, VPA is also used to treat other neurological conditions, including migraine headache and bipolar disorders [330]. Furthermore, it was recently discovered that VPA acts as a positive modulator of chemotherapy in a wide range of cancer treatments [309, 47]. Studies have suggested several mechanisms by which VPA controls seizure generation and propagation. These mechanisms include increasing Gamma-aminobutyric acid (GABA) levels in the cerebral, blocking voltage

dependant sodium and T-type calcium channels, increasing GABA synthesis and release in substantia nigra, reducing neuronal excitation by stimulating N-metyl-D-aspartate (NMDA) glutamate receptors [31, 259, 98, 22]. Furthermore, VPA was observed to activate the extracellular signal-regulated kinase (ERK) pathway, which is involved in neurogenesis and neuronal plasticity [31, 259]. Besides its role in neuronal membrane excitability, VPA is an established inhibitor of histone deacetylase (HDAC), a negative regulator of gene expression [22].

Although VPA is considered the first-line treatment for epileptic children and adults with either focal or general seizures [203], recent evidence suggests that VPA exerts both long-term beneficial and detrimental effects on neurons [49, 243, 320]. Neurodevelopmental Effects of Antiepileptic Drugs (NEAD) study reported cognitive impairments, including verbal and non-verbal abilities, memory, IQ and executive function in children of mothers treated with VPA [212]. In accordance with the human data, similar results were observed in animal model studies in which a higher incidence of autism was observed in the offspring born to mothers taking VPA during their first trimester of pregnancy [261]. On the contrary, positive cognitive effects were also observed in animal models of epilepsy [38, 201, 115]. Furthermore, animal-based studies have demonstrated that early and longitudinally maintained VPA exposure caused dendritic morphology alterations [289]. So far, it has been suggested that ASD like behaviour observed in VPA treated experimental models of epilepsy may be due to an imbalance between GABAergic and glutamatergic transmission; therefore, studies have proposed the use of an endogenous NMDA receptor antagonist, agmatine, to rescue activated ERK1/2 signalling in the brain thus restoring VPA-induced neurobehavioral impairment [156].

4.1.3 Translational animal model of VPA-induced teratogenicity

Although animal studies have been successfully used to give insight into the mechanisms by which VPA induces gene expression changes in the central nervous system and neurons [49, 63] , their relevance to human neurodevelopmental outcomes remains poorly defined. This could be due to the use of non-clinically utilised drug delivery systems or the use of non-therapeutically relevant VPA concentrations observed in human subjects. Recently, Jazayeri et al. (2020)

[135] developed and validated a translational animal model of VPA induced teratogenicity that mimics several clinical features of human gestational VPA exposure. Furthermore, the authors managed to treat maternal rats before and after conception with VPA oral treatment concentration reflective of VPA blood levels observed in humans [135]. The model recapitulated VPA-induced brain abnormalities, including maldevelopment, altered intravertebral distances and a significant developmental delay of vertebral arches. Therefore, the model evaluated the molecular mechanisms underpinning VPA-induced adverse neurodevelopmental outcomes.

4.1.4 Aims and objectives

The main aim of this study were:

1. To evaluate the mechanistic underpinnings of VPA induced neurodevelopmental effects using brain expression changes observed in rat pups chronically exposed to VPA *in utero*.

2. To investigate the influence of epilepsy and genetic background on gene expression changes (for this both epileptic (Genetic Absence Epilepsy Rats from Strasbourg (GAERS), a model of GGE) and Non-Epileptic Control (NEC) dams were used).

4.2 Methods

4.2.1 Experimental Design

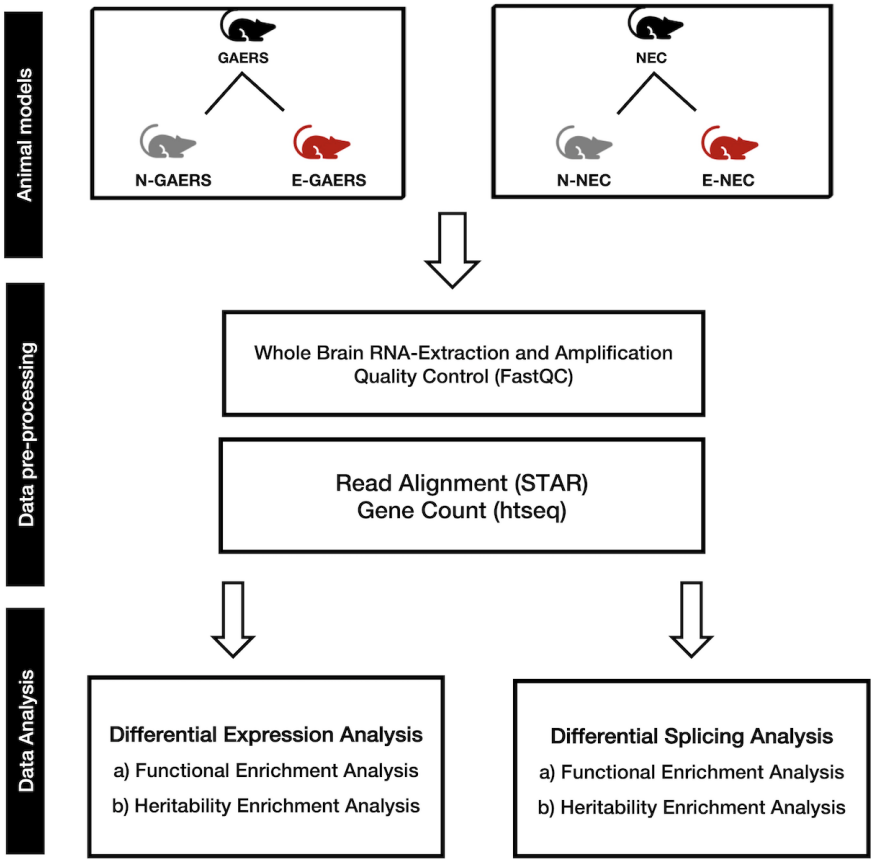


Figure 4.1: **Overview of experimental design and data processing.** Two genetic strains of rats were studied, Genetic Absence Epilepsy Rats from Strasbourg (GAERS) rats and Non-Epileptic (NEC) control rats. Whole brain samples were extracted from valproate-exposed (E-GAERS and E-NEC) and non-exposed (N-GAERS and N-NEC) pups (total n=30). Gene expression was assayed using RNA-Sequencing (RNA-Seq). Data pre-processing steps included quality control for sequence reads alignment to the rat genome and library size normalization. Differential gene expression and differential splicing analyses were carried out with downstream analyses consisting of pathway and enrichment of heritability analyses. (Feleke et al. (2022))

All wet lab-based work, including animal selection, was performed by Professor Terence J. O'Brien group members. Two strains of female rats, inbred NEC and GAERS, were obtained from the Department of Medicine, Royal Melbourne Hospital, University of Melbourne Biological Research Facility [48]. VPA was administered orally at a clinically relevant dose. The rodents were fed either standard chow or a standard rodent diet mixed with 20g/kg VPA [135]. The presence of a plug was used to mark Day 0 of pregnancy. Foetuses were extracted from the

uterus on the 21st day of the pregnancy using Caesarean section (C-section). Thereafter, the QIAGEN Simultaneous Purification of Genomic DNA was used to extract pup brain tissues and immediately samples were frozen and stored at -80 degree Celsius. Additionally, foetuses were examined for any form of birth defects using external assessments including spinal measurements (Figure 4.1).

4.2.2 Transcriptome sequencing and gene expression data processing

Total RNA was extracted from pup brain tissue samples using mRNA Isolation Kit, following manufacturer's instructions and the quality and quantity of the RNA was examined using the TruSeq Stranded Total RNA Library Prep Gold. Sequencing libraries were quantified and sequence reads per sample generated. FastQC [10] was used to carry out QC of RNA-Seq reads. Low-quality reads and ribosome RNA reads were removed and final trimmed reads were mapped to the Rattus norvegicus reference genome (rno6) using STAR [75] (v 2.5).

Fetal sex determination

Strain	VPA-Exposure	Number of pups	Abbreviation	Male	Female
GAERS	Exposed	7	E-GAERS	4	3
	Non-Exposed	8	N-GAERS	3	5
NEC	Exposed	7	E-NEC	4	3
	Non-Exposed	8	N-NEC	5	3

Table 4.1: **Number of pups per group and exposure status.** Abbreviations: GAERS (Generalised epileptic rats from Strasbourg); VPA (Sodium Valproate).

Fetal sex determination was carried out using a list of prenatally expressed Y-chromosome-encoding genes derived from a recently published study [253]. The list of genes was generated by comparing mRNA levels in different brain regions from human prenatal brain samples. In the current study, the expression of marker genes were assessed using the rodent orthologs:

Ubiquitin Specific Peptidase 9 Y-Linked (USP9Y), DEAD-Box Helicase 3 Y-Linked (DDX3Y), and Ubiquitously Transcribed Tetratricopeptide Repeat Containing, Y-Linked (UTY). The samples were then categorised into male or female based on the expression profile of those genes. In total,14 samples were categorized as female and 16 as male (Table 4.1).

4.2.3 Differential gene expression analysis

Differential expression analysis was performed using the R package `edgeR` [257]. A threshold of FDR < 0.05 was used to identify DEGs. After regressing any technical differences due to sex, mRNA expression differences were explored between VPA-exposed and non-exposed pup brains. The comparisons were done in the following order:

1. Exposed Non-Epileptic Control (E-NEC) vs Non-Exposed Non-Epileptic Control (N-NEC)

2. Exposed Genetic Absence Epilepsy Rats from Strasbourg (E-GAERS) vs Non-Exposed Genetic Absence Epilepsy Rats from Strasbourg (N-GAERS)

3. All exposed pups (i.e., E-GAERS + E-NEC) vs All non-exposed pups (N-GAERS + N-NEC)

Although the primary aim of the study was to assess transcriptional changes following gestational VPA exposure, the gene expression changes in the rat pup brain due to epilepsy were also conducted in the following order:

1. E-GAERS vs E-NEC

2. N-GAERS vs N-NEC

(Appendix C, Table 1)

4.2.4 Functional enrichment analysis

Functional enrichment analyses of DEGs from each pairwise comparisons were performed using the R package `WebGestaltR` (v 0.4.4) [182], which uses databases from Gene Ontology Consortium. Default values for `WebGestaltR` default parameters, including the overlap of 10 for minimum and maximum 500 genes. P-values were corrected (for multiple testing) using the BH FDR-correction method. Significant pathways were then identified using a threshold of FDR < 0.05. Significant GO-terms were reduced in the following order:

1. To avoid redundancy, GO terms with ≥ 20 genes or ≤ 1000 genes were filtered out for each of the analyses.

2. `go_reduce()` function from the R package `r-utils` was used to calculate semantic similarity of GO terms.

3. A threshold of 0.9 was applied to the hierarchical tree generated by `reduceSimMatric()` function from the R package `rrvgo` (v 1.1.4) (which uses a bottom-up clustering method) to get fewer GO-terms. (Appendix C, Table 2a-c).

4.2.5 LD regression

LDSC was used to test if the dysregulated gene in the pups' brain due to VPA exposure were enriched for common genetic variants associated with 7 neurological traits, namely ADHD [69], Bipolar Disorder (BD) [293], ASD [105], schizophrenia (SCZ) [255], full-scale IQ [272], Epilepsy (EPI) [1] and Cross Disorders Group (CDG) which is a GWAS meta-analysis across 8 psychiatric traits [174]. The following steps were taken to test for enrichment of heritability of the DEGs for any of the 7 traits:

1. Single nucleotide polymorphisms (SNPs) annotation files containing rows corresponding to a SNP and a column for each sub-annotation were generated and dsSNP file NCBI Build 37 co-ordinates (build 147 and hg19) were used to map the SNPs to genes and values of 0 were given to SNPs not present in the file. Ten annotation files were generated

for all the comparisons, 5 files were for up-regulated and 5 were for down-regulated DEGs from each comparison.

2. For each generated file, LDSC was run using data files from phase 3 of the 1000 Genome Project Phase 3 European population. LD scores were calculated for the annotations using 1cM window (a default value), and the analysis was restricted to Hapmap3 SNPs.

3. LDSC python scripts `munge_sumstats.py` were used to format the summary statistic files, and for the regression weights, LD calculated for HapMap3 SNPs were downloaded from the LDSC Github page (`https://github.com/bulik/ldsc`) – the full baseline model was used for the analysis.

4. An enrichment score and its associated P-value was calculated based on the proportion of total SNPs per annotation (column), after considering all other annotation. Annotation categories with significant positive (FDR < 0.05) enrichment of SNP-heritability were then reported as a final result (a one-tailed test).

GWAS summary statistics were obtained from The European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) and Psychiatric Genomic Consortium (PGC) Cross-Disorder Group. All subjects were of European ancestry. A full detailed list of all the summary statistics used in these analyses can be found in Appendix C, Table 3.

4.2.6 Deconvolution

To deconvolute the bulk RNA-seq signature into its component single cell-type expression profiles, a scRNA-Seq dataset derived from Manno et al., 2021 [94] was utilised. In the study, the authors constructed a single-cell transcriptomic atlas of the embryonic mouse brain from gastrulation up until birth [164]. In the current study, cells from embryonic day 18 (E18) were extracted, and gene by cell-matrix was generated. Following this, cell-types were identified using functions from the R package, `Seurat` [270]. Cell-type identification was carried out in the following order:

1. The dataset was normalised using the `NormalizeData()` function, which uses a global scaling normalisation method that normalises the gene expression values by multiplying the total expression of the cell by the number (n) of cells and log transforms the values for a final expression values (a default value of 10,000 was used as a scaling factor).

2. The distances between two cells with similar expressions were calculated using the Euclidean algorithm, and edges were drawn.

3. Cells were clustered using `FindClusters()` function (parameters used: 30 PCs and 2 resolution value).

4. The clusters were visualised using non-linear dimensionality reduction algorithm UMAP (v 0.1.10).

5. DEGs in one cluster compared with all other clusters (cluster-specific genes) were identified using the function `FindAllMarkers()` (DE method: rank sum test (FDR < 0.05)).

6. Cell-types were allocated by testing if a cluster-specific DEGs set was enriched (Fisher's exact test) for cell-type specific marker genes from a gene list derived from Zeisel et al.(2018) [356].

Once cell-types were identified for each cluster, the relative proportions of cell-types in rat pups brain tissue samples were determined using, weighted least squares-based deconvolution algorithm Dampened Weighted Least Squares (DWLS). The following steps were taken to generate relative proportions of cell-types per group, namely, E-NEC, E-GAERS, N-GAERS, and N-NEC:

1. Expression matrix from samples was normalised using counts per million based normalisation methods.

2. `biomaRt` R package (v 4.1)[79] was used to infer the orthologous mouse gene (for the rat genes) (Ensembl Genes 104).

3. Cells were deconvolved using weighted least squares approach. (`https://github.com/dtsoucas/DWLS`).

4.2.7 Code and Data availability

The raw and processed rat sequencing data generated in this study have been deposited in NCBI Gene Expression Omnibus database under accession number GSE198756. All scripts used in this study can be found in GitHub `https://github.com/rahfel/VPA`. Supplementary materials including full tables can also be found in GitHub `https://github.com/rahfel/VPA)`.

4.3 Results

4.3.1 Study workflow and data collection

Female rats were taken from Wistar rat colony and selectively inbred to either express or not express absence seizures (a summary of the study workflow can be found in Figure 4.1) [48, 198]. The rats were maintained at a facility on a light-dark cycle of 12 hours light, 12 hours dark and at a temperature range of 19 to 22 degree Celsius. The rats were fed a diet consisting of either a pre-mixed meal or a standard rodent diet as listed by Senn and colleagues [274] for 2 weeks before mating with males of the same strain and continued throughout pregnancy. This dose of orally administrated VPA was initially reported in the study by Jazayeri et al. (2020) [135], and resulted in a significant seizure suppression in adult GAERS.

Moreover, measured blood serum levels were equivalent to human therapeutic levels of VPA. Once the plug was present (day 0 of pregnancy), females were separated from the males for the course of their pregnancy, and the following day was classified as the first day of pregnancy. The fetuses were extracted from the uterus via C-section one day before expected birth (day 21), and the fetus's brains were removed, snap frozen and stored at -80 degree Celsius. The teratology studies, including spinal measurements are reported in Jazayeri et al. (2020) [135].

In total, 7 litters were generated for each treatment group and pup brains were randomly processed from each litter (one for exposed groups and one or two for non-exposed groups). Total RNA was extracted from 30 pup brains (Table 4.1) and genome-wide gene expression data was generated using RNA-Seq.

4.3.2 Deconvolution analysis reveals no difference in cell-type composition between VPA-exposed and non-exposed pup brains

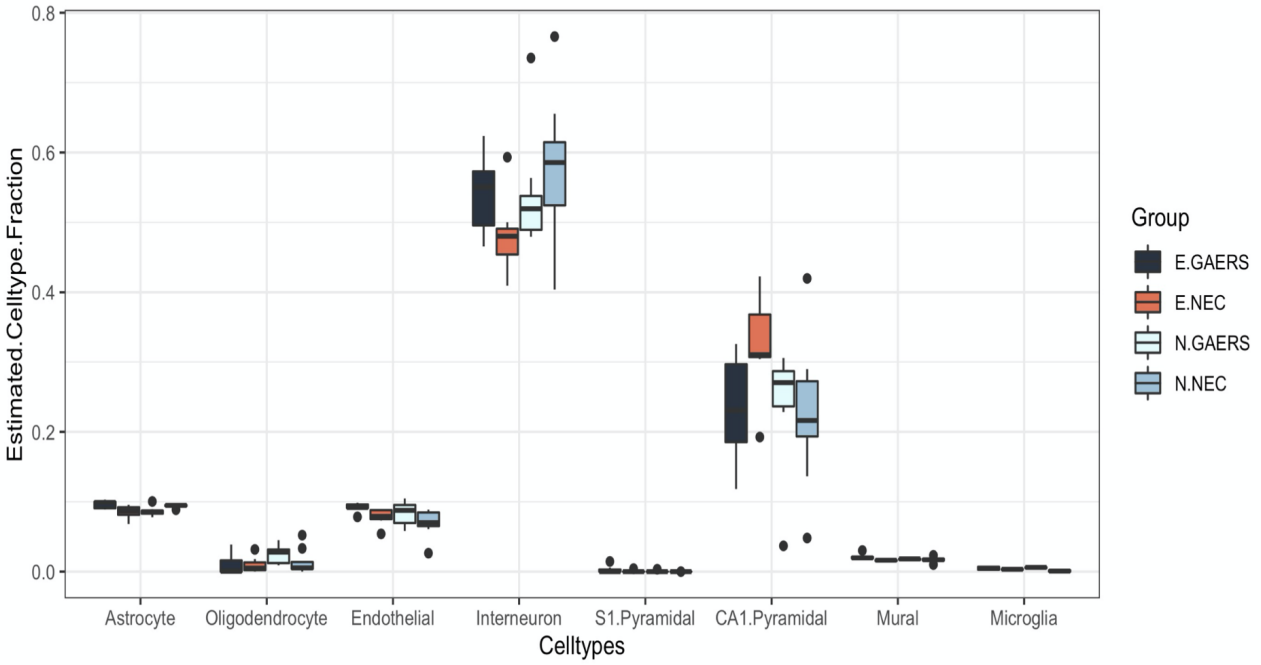


Figure 4.2: **Deconvolution on bulk samples showing cell-type proportion for each group comparison and for each major brain cell-type.** Wilcoxon rank sum test was used to determine the significance of differences in cell-type proportion between conditions. FDR correction for multiple testing was applied. There were no significant differences in cell-type proportion at FDR < 5 for any of the valproate-exposured vs non-exposured comparisons. Refer to Appendix C Table 4 for DWLS cell fractions and FDR values.

Cell-type proportions in bulk-tissue RNA-Sequencing samples were estimated to test whether *in utero* VPA exposure is associated with significant changes in the proportion of cell-types in the rat pup brain. A weighted least squares-based deconvolution algorithm, DWLS, (which estimates cell-type composition in a bulk tissue RNA-seq dataset using prior information from an unrelated scRNA-Seq signature from an analogous tissue) was used to assess the cell-type proportional changes due to VPA exposure. Taking this approach, there was no significant (FDR-corrected Wilcoxon rank sum test) difference in cell-type composition observed between valproate-exposed and non-exposed pup brains, for either GAERS or NEC rats, suggesting gestational VPA exposure does not cause apoptosis and alter gene expression via epigenetic mechanisms (Figure 4.2 and Appendix C, Table 5).

4.3.3 Differential gene expression analysis highlights transcriptional changes in VPA-exposed pup brains

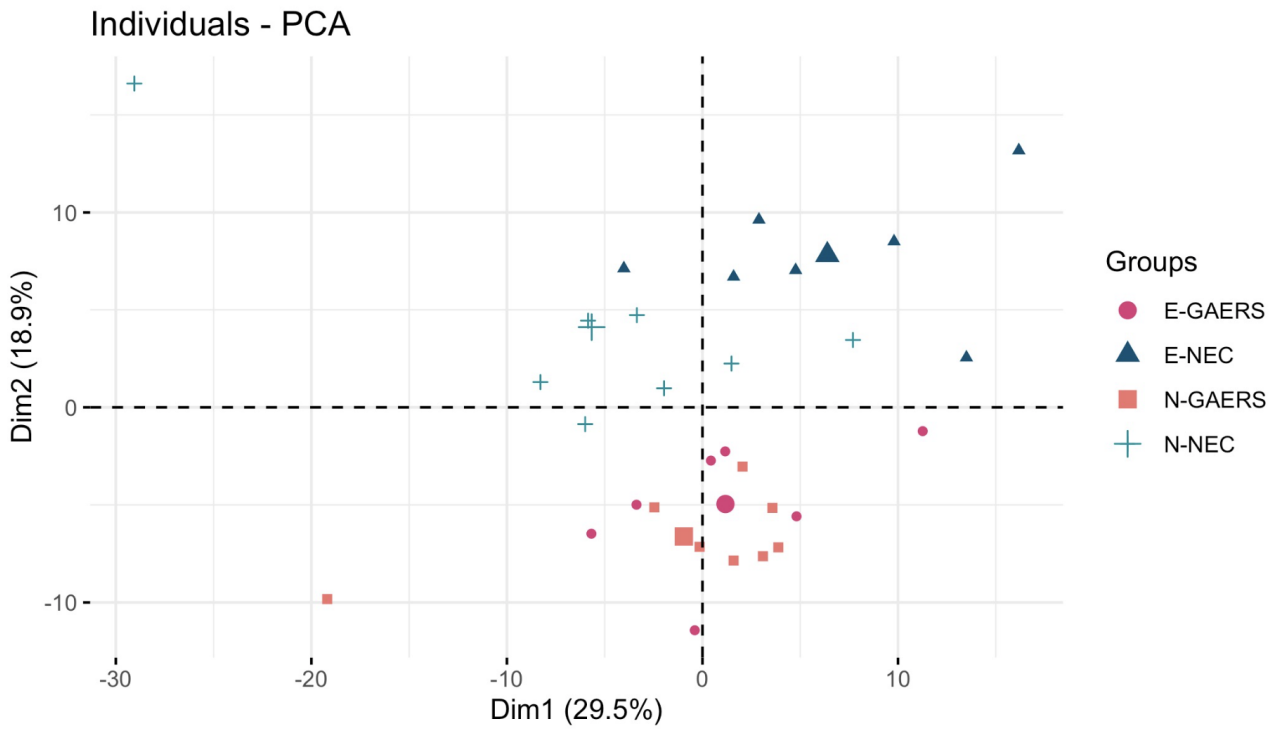


Figure 4.3: **Principal component analysis plots.** Epileptic and non-epileptic rat pups are broadly separated by PC1 and VPA-exposed and non-exposed pups by PC2.

The principal components of variation between the groups were explored prior to carrying out differential expression analysis. The function `prcomp()` from the R package `stats` (v 3.6.2) was used to calculate the PCs, and the function `fviz_pca_ind()` from the R package `factoextra`(v 1.0.7) was used to visualise the PCs (Figure 4.3). The epileptic and non-epileptic pups were separated by the first principal component (PC1) and VPA-exposed and non-exposed pups were separated by PC2. Following this, differential gene expression analyses for case (exposed) versus non-exposed comparisons were undertaken (a full list of comparisons carried out is found in Table 4.2).

Focusing on VPA-exposed versus non-exposed pup brains, the largest number of DEGs (FDR $<$ 0.05) were observed when comparing all VPA-exposed pups (i.e., All Exposed (All-E) consisting of E-GAERS + E-NEC pups) versus all non- exposed pups (i.e., All Non-Exposed (All-N) consisting of N-GAERS + N-NEC pups), where 3,470 genes were significantly (FDR $<$ 0.05)

Comparisons	Sample size	Number of DEGs (FDR <0.05)	Up-regulated (FDR <0.05)	Down-regulated (FDR <0.05)
E-NEC vs N-NEC	15	2,973	1,307	1,666
E-GAERS vs N-GAERS	15	553	248	305
All-E (E-GAERS + E-NEC) vs All-N (N-GAERS + N-NEC)	30	3,470	1,632	1,838
E-GAERS vs E-NEC	14	1,885	1,039	846
N-GAERS vs N-NEC	16	1,066	548	518

Table 4.2: **Differential gene expression analysis: case control comparisons.** Table Abbreviations NEC: non-epileptic control; GAERS: Genetic Absence Epilepsy Rats from Strasbourg; E-NEC: VPA-exposed NEC pups; E-GAERS: VPA-exposed GAERS pups; N-NEC: non-exposed NEC pups; N-GAERS: non-exposed GAERS pups; All-E: All exposed pups (E-NEC + E-GAERS); All-N: All non-exposed pups (N-NEC + N-GAERS).

differentially expressed. Of these, 1,632 genes were up-regulated and 1,838 down-regulated. Figure 4.4 shows the overlaps in genes DE for each of the 3 pairwise VPA-exposed versus non-exposed comparisons (i.e., E-NEC versus N-NEC, E-GAERS versus N-GAERS and ALL-E versus ALL-N). Refer to Appendix C, Table 1 for a full list of genes DE in each comparison.

Moreover, there was a substantial overlap between genes DE following VPA exposure in GAERS pups compared to genes DE in NEC pups. Further, the majority of genes differentially expressed by VPA in GAERS pups were also differentially expressed in NEC pups, suggesting that genetic epilepsy status is not a major determinant of the pattern of differential expression induced by VPA, as was previously the case for VPA induced birth defects in this model [135] (Figure 4.4).

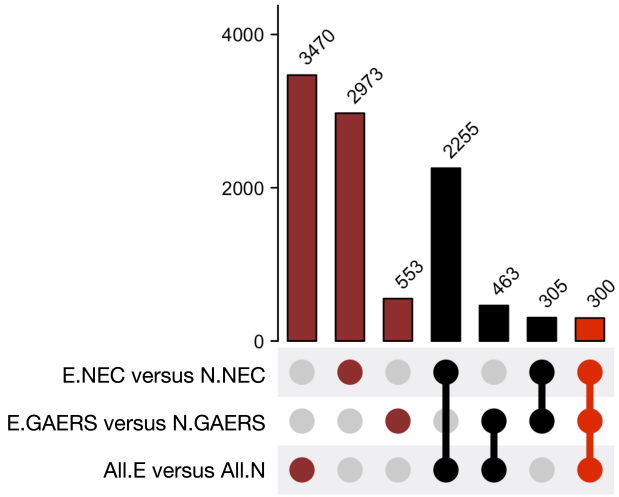


Figure 4.4: **Summary of differential gene expression results for valproate-exposed versus non-exposed comparisons**. Total number of differentially expressed genes (DEGs) for each comparison are shown (single brown dot corresponding to the first three histograms). The number of genes overlapping between each of the case versus control comparisons are shown in black (2-way) or red (3-way). The bars are arranged based on the highest number of overlaps. (Feleke et al. (2022))[86]

4.3.4 Pathway enrichment analysis reveals functional consequences of VPA-Induced differential gene expression in the developing brain

Pathway enrichment analyses of DEGs were conducted to assess the functional consequences of VPA induced differential gene expression in the developing brain. Interestingly, genes that were down-regulated following VPA exposure were significantly enriched for functional processes relating to modulation of synaptic function and neuronal processes (Figure 4.5; refer to Appendix C, Tables 2 for full details relating to pathways enrichment analysis, including significantly enriched biological processes, cellular components and molecular functions GO terms).

When considering all VPA exposed pups versus non-exposed pups (All-E vs All-N), among the set of genes down-regulated by VPA were found to be significantly enriched for terms relating to glutamate receptor complex (FDR $= 9.74 \times 10^{-12}$), regulation of membrane potential (FDR =

2.08×10^{-14}), synapse assembly (FDR = 2.17×10^{-11}), regulation of actin cytoskeleton organisation (FDR = 9.94×10^{-05}), post-synaptic membrane (FDR = 2.11×10^{-04}), neurotransmitter receptor activity (FDR = 2.33×10^{-04}) and axon guidance (FDR = 1.97×10^{-12}). Furthermore, among the pathways enriched in the down-regulated set of genes was the regulation of insulin secretion (FDR = 9.5×10^{-03}), suggesting a possible drug-induced transcriptional mechanism for the increased incidence of impaired glucose control in patients treated with VPA [14, 281].

In contrast to the substantial enrichment of neuronal functions in genes down-regulated by VPA, the genes up-regulated by VPA were generally enriched for functional terms not directly related to neural processes namely mRNA splicing (FDR = 7.29×10^{-14}), translation (FDR = 1.40×10^{-13}), extra-cellular matrix organisation (FDR = 2.41×10^{-12}) and cell cycle (FDR = 2.40×10^{-09}). Furthermore, there was an enrichment for biological processes relating to chromatin organization (FDR = 7.37×10^{-07}) and chromatin assembly/disassembly (FDR = 1.45×10^{-06}) among the genes up-regulated by VPA. Given the functional enrichment of chromatin assembly/disassembly terms among the genes upregulated by VPA, along with VPA's known activity as a histone deacetylase inhibitor [22], it seems likely that this transcriptional dysregulation is epigenetically encoded with potentially long-lasting consequences for the human brain function and health even in the absence of valproate exposure.

Taken together, genes up- and down-regulated by chronic *in utero* VPA exposure in the developing rat brain were represented by divergent functional categories, including neuronal and non-neuronal pathways. Additionally, genes down-regulated by VPA were predominantly characterized by pathways relating to nervous system function.

Therefore, the relationship between genes differentially expressed in pup brains gestationally exposed to VPA and genetic determinants of neurodevelopmental disease and behavioural traits were investigated. Further, given that neurodevelopmental diseases were shown to arise from genetic mutations, namely deleterious and loss-of-function mutations [238], in rare variant genetic analyses, it was hypothesised that genes significantly down-regulated by VPA in the developing brain are enriched for genetic risks of neurodevelopmental diseases.

4.3.5 Genes down-regulated by gestational VPA exposure are highly enriched for heritability for neurodevelopmental traits and disease

In the context of rare variant genetic analyses, neurodevelopmental disability primarily arises from loss of function and deleterious mutations [142]. Therefore, here LDSC [40] method was used to investigate if genetic risks for neurodevelopmental diseases and behavioural traits are enriched in the set of genes significantly down-regulated in pup brains following gestational VPA exposure. Given the broad range of adverse neurodevelopmental outcomes in children gestationally exposed to VPA, the following brain disorders and neurobehavioral traits related GWAS were considered: (i) ADHD [69]; (ii) BD [293]; (iii) ASD [105]; (iv) SCZ [255]; (v) full-scale IQ [272]; (vi) EPI [1]; and (vii) CDG [174] which is a GWAS meta-analysis across 8 psychiatric traits. Further, a GWAS study forWaist-to-hips ratio (WHR) [250], which was not expected to be enriched in the set of genes differentially expressed in pup brains as a result of VPA exposure, was used as a negative control.

Enrichment of genetic association analyses were run for all sets of genes significantly (FDR < 0.05) DE in the VPA-exposed pup brain (Figure 4.6; refer to Appendix C, Table 3 for full details). Interestingly, for genes down-regulated by VPA, there was a significant enrichment of heritability for BD (All-E vs All-N, $\text{FDR}_{\text{LDSC}} = 1.16 \times 10^{-08}$; E-NEC vs N-NEC, $\text{FDR}_{\text{LDSC}} = 2.93 \times 10^{-09}$; E-GAERS vs N-GAERS, $\text{FDR}_{\text{LDSC}} = 1.55 \times 10^{-03}$), SCZ (All-E vs All-N, $\text{FDR}_{\text{LDSC}} = 6.04 \times 10^{-08}$; E-NEC vs N-NEC, $\text{FDR}_{\text{LDSC}} = 6.40 \times 10^{-08}$; E-GAERS vs N-GAERS, $\text{FDR}_{\text{LDSC}} = 2.91 \times 10^{-02}$), IQ (All-E vs All-N, $\text{FDR}_{\text{LDSC}} = 1.17 \times 10^{-03}$; E-NEC vs N-NEC, $\text{FDR}_{\text{LDSC}} = 7.03 \times 10^{-04}$; E-GAERS vs N-GAERS, $\text{FDR}_{\text{LDSC}} = 1.62 \times 10^{-02}$), and CDG (All-E vs All-N, $\text{FDR}_{\text{LDSC}} = 3.89 \times 10^{-04}$; E-NEC vs N-NEC, $\text{FDR}_{\text{LDSC}} = 1.82 \times 10^{-04}$; E-GAERS vs N-GAERS, $\text{FDR}_{\text{LDSC}} = 5.44 \times 10^{-02}$). However, no significant enrichments of heritability were observed for these traits in non-exposed GAERS pups compared to non-exposed non-epileptic controls, and enrichments of association for WHR were non-significant across all comparisons (Appendix C, Table 3).

Furthermore, for VPA-exposed pups, there was no significant enrichment of heritability to ADHD (All-E vs All-N, $FDR_{LDSC} = 0.69$; E-NEC vs N-NEC, $FDR_{LDSC} = 0.24$; E-GAERS vs N-GAERS, $FDR_{LDSC} = 0.88$), ASD (All-E vs All-N, $FDR_{LDSC} = 0.66$; E-NEC vs N-NEC, $FDR_{LDSC} = 0.68$; E-GAERS vs N-GAERS, $FDR_{LDSC} = 0.68$). Moreover, there was no significant enrichment of heritability for Epilepsy (EPI) when considering DEG from the All-E vs ALL-N comparison (All-E vs All-N, $FDR_{LDSC} = 0.77$), and for the E-NEC vs N-NEC or E-GAERS vs N-GAERS (E-NEC vs N-NEC, $FDR_{LDSC} = 0.57$; E-GAERS vs N-GAERS, $FDR_{LDSC} = 0.77$). Notably, for genes up-regulated by gestational VPA, there was no significant enrichment of genetic association to any neurodevelopmental disease or trait, other than a marginal enrichment to BD (All-E vs All-N, $FDR_{LDSC} = 2.7 \times 10^{-02}$; E-NEC vs N-NEC, $FDR_{LDSC} = 3.6 \times 10^{-03}$; E-GAERS vs N-GAERS, $FDR_{LDSC} = 0.84$) as well as SCZ (All-E vs All-N, $FDR_{LDSC} = 4.3 \times 10^{-02}$; E-NEC vs N-NEC, $FDR_{LDSC} = 0.32$; E-GAERS vs N-GAERS, $FDR_{LDSC} = 0.56$) concordant with the pathway enrichment analysis.

In summary, these results suggest VPA exerts its adverse effects on fetal neurodevelopment predominantly via the down-regulation of genes highly relevant to neurodevelopment and nervous system function. The directionality of the effect is consistent with that observed from rare-variant analyses of genetic risk to neurodevelopmental disease where the predominant mechanism is a dominant negative effect [142].

4.3.6 Alternatively spliced genes in VPA exposed pups compared to non-exposed pups

Among the pathways enriched in genes up-regulated by VPA, there was a significant (FDR $<$ 0.05) enrichment for genes involved in mRNA splicing related pathways. Therefore, the role of VPA on differential splicing was assessed by first quantifying read count to individual gene exons and then comparing differential exon usage using the R package `edgeR` [257]. In total, there were 57 significantly (FDR $<$ 0.05) alternatively spliced genes, among which 21 were also differentially expressed following VPA exposure (14 were down-regulated, 7 up-regulated) (refer to Appendix C, Table 5 for full result table). Whilst there were no significant enrichments

for known functional pathways or enrichment of genetic association to neurodevelopmental disease among these genes, individual differentially spliced genes included neuronal proteins such as Calmodulin Binding Transcription Activator 1 (CAMATA1) (FDR = 3.69×10^{-26}) which is known to play a role in the regulation of glutamate levels and neuronal excitability, glutamate decarboxylase 2 (GAD2) (FDR = 9.03×10^{-25}) which plays a role GABA-synthesis in neurons [331] and Forkhead box P4 (FOXP4) (FDR = $4.62 \times 10-3$) which is known to regulate neurogenesis and in which mutations are associated with speech delay and congenital abnormalities [262].

4.3.7 A substantial overlap between genes down-regulated in VPA-exposed pup brains and genes genetically linked to increased autism risk

To further assess the association between VPA exposure and ASD risk, the overlap of DEGs taken from VPA-exposed versus non-exposed comparisons and three different ASD-associated gene sets were carried out. A total of 3,162 DEGs (1,454 up-regulated genes and 1,708 down-regulated) remained after converting 3,470 rat genes to their human orthologs. The three sets of ASD-associated genes used in the analysis were:

1. A set of 913 genetically linked ASD genes from Simons Foundation Autism Research Initiative (SFARI) `https://gene.sfari.org/`, an online database of ASD risk genes generated by integrating genetic data from several studies.
2. A set of 1,611 ASD associated genes identified through differential expression analysis performed using transcriptomic data from postmortem brain samples (51 ASD subjects and 936 controls) from the PsychENCODE Consortium [96]
3. A set of 7,418 dysregulated genes in VPA exposed cortical organoids (early stage, day 18) derived from a male iPSC line [63].

In total, 223 VPA-induced DEGs from the current study significantly (P = 9.99×10^{-5}) over-

lapped with genetically linked ASD risk genes from SFARI (Figure 4.7). Strikingly, the overlap was mainly for down-regulated genes (168 of the 223 overlapping genes) in VPA exposed pup brains, suggesting that there is a convergence of molecular and biological pathways underlying ASD and VPA-induced postnatal neurodevelopmental outcomes. Similarly, 219 of the 320 (nominally significant overlap, P = 0.067) overlapping genes between VPA-induced DEGs and genes dysregulated in PsychENCODE ASD brains organoids were down-regulated in the pup brains (Figure fig:Figure 4.7.Overlap). There was no significant (Overlapped genes = 1,366, P = 0.55) overlap between DEGs from the current study and DEGs identified in VPA-exposed organoids from the Cui et al. (2020) [63] (Figure 4.7).

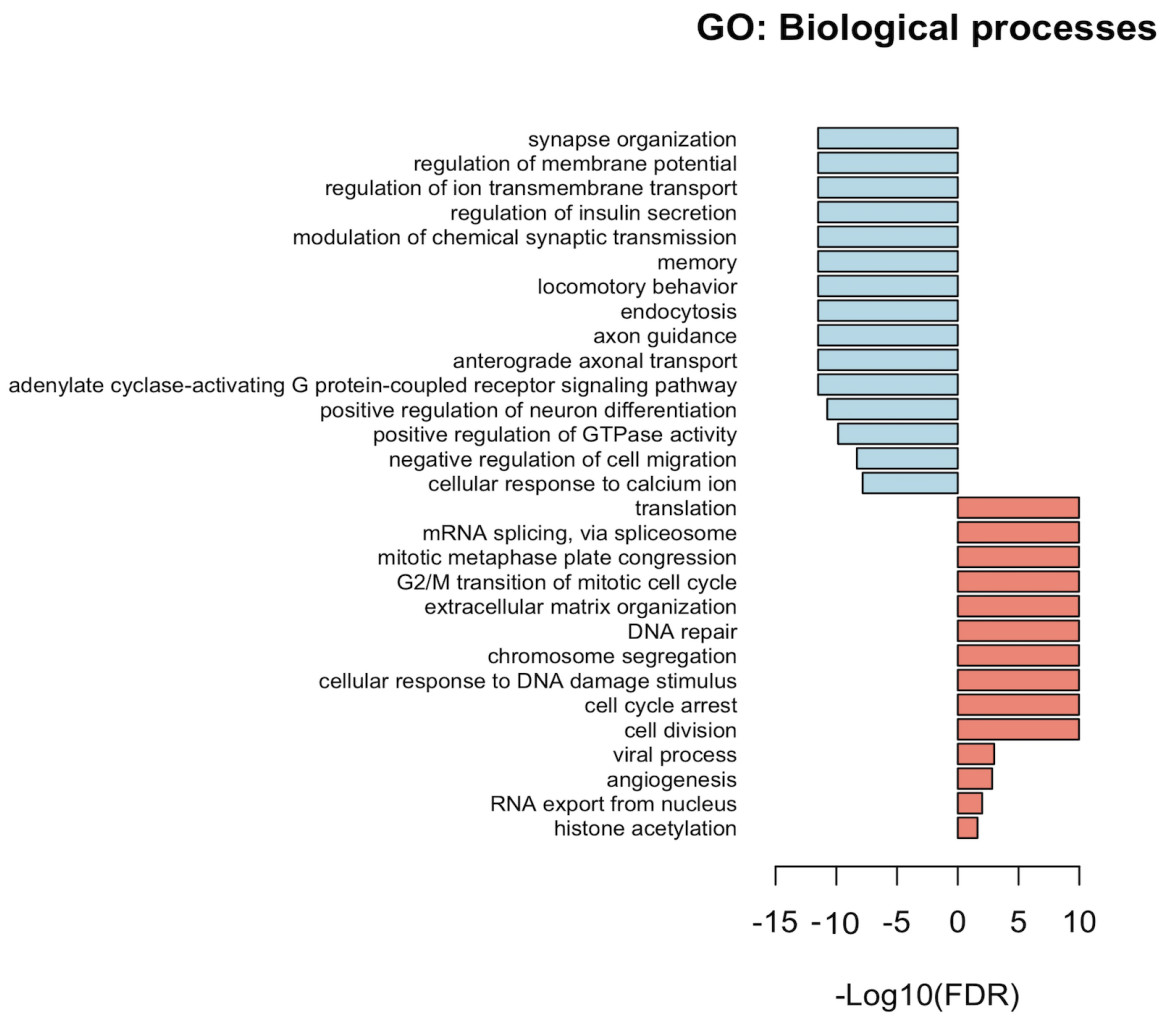


Figure 4.5: **Pathway enrichment for valproate-exposed versus non-exposed pups.** Partitioned Linkage Disequilibrium Score Regression (LDSC) was used to test for enrichment of heritability in genes differentially expressed for each of the 5 case-control comparisons. Enrichment -log10(FDR) for the enrichment of genetic association is indicated by the horizontal bars coloured by the comparison group from which the differentially expressed genes were identified. Vertical line indicates FDR values at 0.05. Genome-wide association studies used for the enrichment analysis are indicated on the vertical axis. Abbreviations: ASD: Autism Spectrum Disorder, ADHD: Attention Deficit Hyperactivity Disorder, BD: Bipolar Disorder, SCZ: Schizophrenia, IQ: Intelligence Quotient, EPI: Epilepsy, WHR: Waist-to-Hip-Ratio, CDG: Cross-Disorder Group. (Feleke et al. (2022))[86]

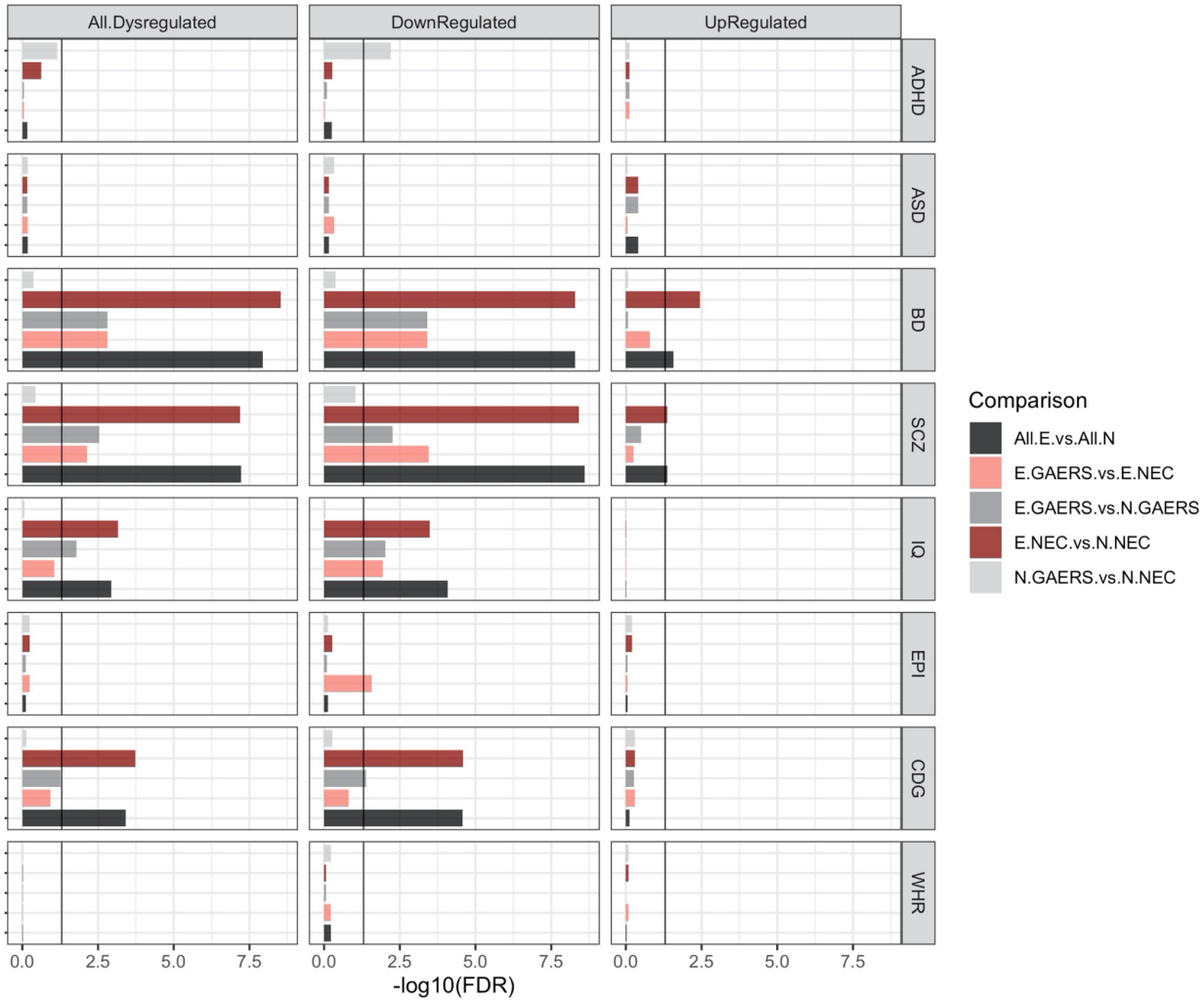


Figure 4.6: **Heritability enrichment in differentially expressed genes.** Partitioned LDSC was used to test for enrichment of heritability in genes differentially expressed for each of the 5 case control comparisons. Enrichment -log10(FDR) for the enrichment of genetic association is indicated by the horizontal bars coloured by the comparison group from which the differentially expressed genes were identified. Vertical line indicates FDR values at 0.05. Genome- wide association studies used for the enrichment analysis are indicated on the vertical axis. Abbreviations: ASD: Autism Spectrum Disorder, ADHD: Attention Deficit Hyperactivity Disorder, BD: Bipolar Disorder, SCZ: Schizophrenia, IQ: Intelligence Quotient, EPI: Epilepsy, LDSC: Linkage Disequilibrium Score Regression, WHR: Waist-to-Hip-Ratio, CDG: Cross-Disorder Group. (Feleke et al., 2022)

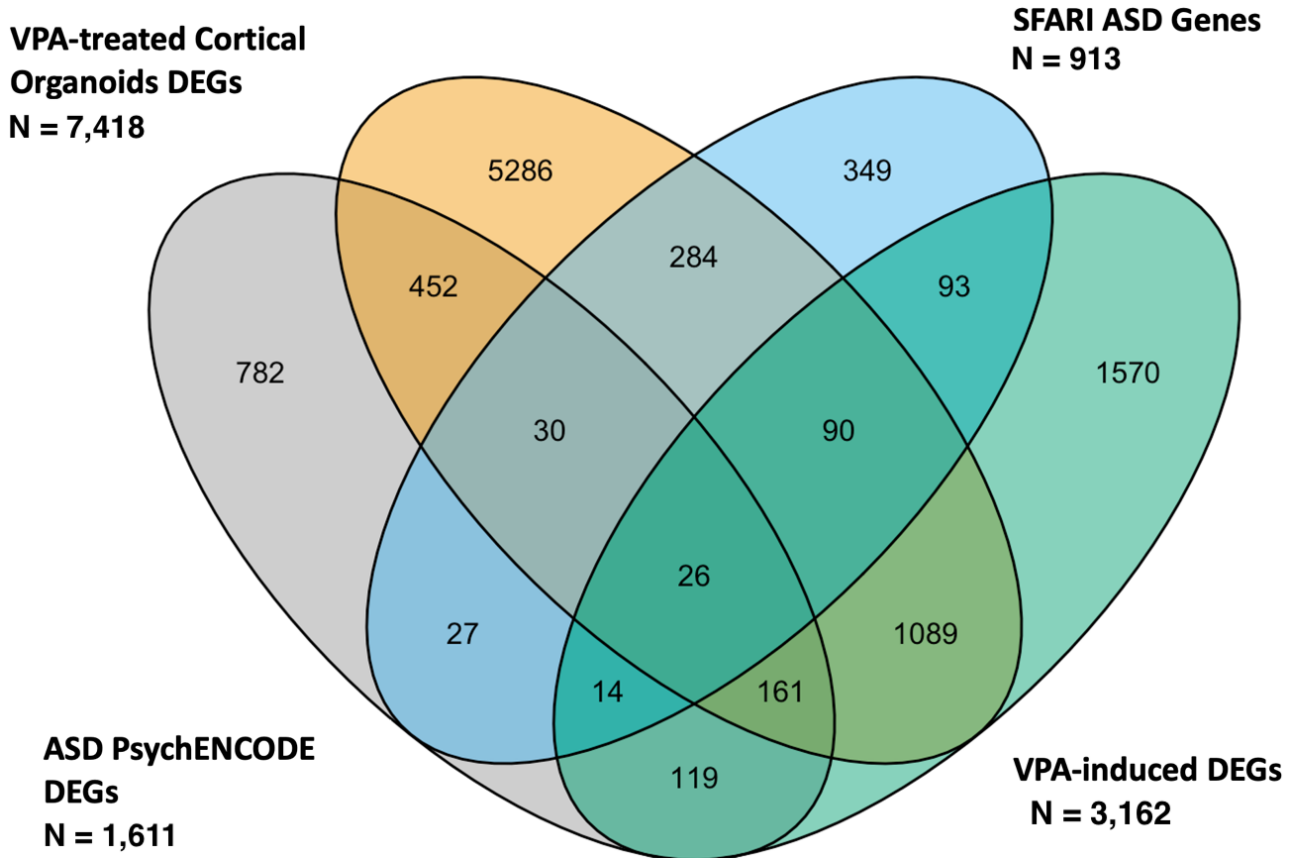


Figure 4.7: **Overlapping genes between different ASD-associated genes.** VPA-induced DEGs set were genes identified in the current study. Abbreviations: ASD: Autism Spectrum Disorder; N: Number of genes in the set, VPA: Sodium Valproate, DEGs: Differentially Expressed Genes, SFARI: Simons Foundation Autism Research Initiative. (Feleke et al.(2022)) [86]

4.4 Discussion

Although VPA is considered one of the most effective treatments for epilepsy [203], women of childbearing potential are advised against the use of the drug due to its well-known and documented adverse neurodevelopmental consequences. Moreover, the biological mechanisms by which valproate contributes to behavioural and cognitive problems in children following gestational exposure remain poorly defined. Further, few studies have examined transcriptome-wide alterations in brain gene expression [63] and often are limited by non-physiological drug administration and dosing regimens. In this study, an established rat model of valproate-induced teratogenicity [135] that recapitulates human prenatal valproate exposure and chronicity of oral dosing during pregnancy was utilized to identify and assess the mechanisms by which VPA exposure induces adverse fetal health effects.

4.4.1 No substantial shifts in the composition of the major cell-types

There was no evidence for substantial shifts in the composition of the major cell-types of the brain (i.e., excitatory neurons, inhibitory neurons, astrocytes, oligodendrocytes, oligodendrocyte precursor cells, microglia) following VPA-exposure using deconvolution analysis. This does not reflect the proportion of cell-type specific sub-types occurring as a result of VPA exposure, which will require single-cell RNA-sequencing of large numbers of cells from VPA-exposed and non-exposed brains; however, these findings are consistent with no measurable differences in the proportion of the major cell-types. Focusing on major cell-types, the observed differences in gene expression between exposed and non-exposed brains arise predominantly from differential gene expression rather than substantial differences in the proportions of the major cell-types between VPA exposed and non-exposed brains.

4.4.2 Transcriptional dysregulation in pup brains following gestational VPA exposure

Furthermore, there was statistically significant (FDR < 0.05) transcriptional dysregulation in the pup brains following gestational VPA exposure. Pathway enrichment analysis revealed a

noticeable separation of enriched terms between the up- and down-regulated genes. While, up-regulated genes were predominantly enriched for terms relating to mRNA splicing division, translation, cell division and extracellular matrix organisation, down-regulated genes were highly enriched for neurodevelopmental-related terms, including Extracellular Signal-Regulated Kinase 1 (ERK1) and Extracellular Signal-Regulated Kinase 2 (ERK2) cascade, regulation of neuronal membrane activity, glutamate receptor signalling pathway, gamma-aminobutyric acid signalling pathway and synaptic transmission.

4.4.3 Enrichments of heritability in the down-regulated genes

Differentially expressed genes were integrated with GWAS summary statistics from a wide range of neurological traits and diseases to assess the functional consequences of VPA-induced differential gene expression. There was significant heritability enrichment in the genes down-regulated by VPA for bipolar disorder (unsurprising, due to the fact that VPA is widely used to treat bipolar disorder), schizophrenia, IQ and cross disorder (Figure 4.6). Though there was no enrichment of heritability for ASD in the down-regulated gene sets (which was surprising given that some of the features of the valproate-neurodevelopment syndrome have been likened to autism, and maternal valproate exposure has been associated with autism-like behaviours in non-human primates) [361], there was a significant ($P = 9.99 \times 10^{-05}$) overlap between genes genetically linked to ASD and down-regulated genes in the VPA exposed pup brains 4.7. These include genes SCN1A, Sodium Voltage-Gated Channel Alpha Subunit 2 (SCN2A) [214], Gamma-Aminobutyric Acid Type A Receptor Subunit Beta3 (GABRB3) [305], Hyperpolarization Activated Cyclic Nu- cleotide Gated Potassium Channel 1 (HCN1) [200], Calcium/Calmodulin Dependent Protein Kinase IV (CAMK4) [355], and ATPase Plasma Membrane Ca2+ Transporting 2 (ATP2B2) [349] which harbour mutations responsible for the increased risk of ASD and other neurodevelopmental disorders. However, due to the absence of a genetic association with ASD (from the enrichment of heritability analyses results Figure 4.6), which was in stark contrast to the enrichment for bipolar disorder, schizophrenia and IQ, it is possible that valproate fetal neurodevelopmental syndrome may have specific clinical characteristics unique to valproate exposure and points to a requirement to continue to define

the phenotype.

It is worth noting that the directions of effect of valproate-induced gene expression on brain function is consistent with that observed from rare-variant analyses of genetic risk to neurodevelopmental disease where the predominant mechanism is a dominant negative effect from deleterious mutations [142]. The functional enrichment of chromatin assembly/disassembly terms among the genes up-regulated by VPA suggests the VPA induced transcriptional dysregulation is epigenetically encoded and, therefore, may have potentially long-lasting consequences for human brain function, even after birth prompting the need for further clinical research on the long-term outcomes of children born following fetal VPA exposure.

4.4.4 Differential mRNA splicing in the brain

Given the significant (FDR = 7.29×10^{-14}) enrichment for genes involved in mRNA splicing in functional enrichment analysis, differential splicing analysis was performed to assess if VPA-induced transcriptional dysregulation is epigenetically controlled. In addition, recent studies have shown evidence implicating the regulatory role of neuron-specific alternative splicing in neurodevelopmental disorders [247, 334]. Further, alternative splicing in the brain is vital for several neurological processes, including cell differentiation, neurogenesis, synaptogenesis and the generation of functional neuronal networks [300]. Of those genes observed to be significantly differentially spliced in the prenatal brain following VPA exposure were GAD2 (also known as GAD65), which plays a role in GABA-synthesis in neurons [331] and FOXP4, which is known to regulate neurogenesis and is associated with speech delay and congenital abnormalities [262]. These results suggest that alternative splicing may be an additional mechanism for adverse neurodevelopment in VPA-exposed fetal brains.

4.4.5 Other anti-epileptic drugs

Though, valproate is the most widely recognised AED to affect neurodevelopment and significantly increase the risk of ASD, prenatal exposure to other anti-seizure medications such as

phenytoin and phenobarbital was also associated with an increased risk of neurodevelopmental deficits [152]. However, this was not observed for lamotrigine or levetiracetam, which appear safer with respect to cognitive and behavioural outcomes [330, 152]. Additionally, studies have suggested that anti-depressant use during pregnancy may be associated with an increased risk of neurodevelopmental disorders, including ASD and ADHD [29, 335], although not for exposure to antipsychotics [335]. The research presented here, which demonstrates a robust brain transcriptional response to VPA that is both functionally and genetically associated with relevant cognitive (IQ) and psychiatric (BD, SCZ, CDG), outcomes suggests that the rat model of chronic dosing followed by transcriptional assay in pup brains may provide a more general approach for screening for drug-induced adverse neurodevelopmental effects. Moreover, the inference that the adverse behavioural and cognitive outcomes from gestational VPA exposure arise from transcriptional dysregulation, the model presents a potential system for testing drugs capable of reversing or improving these changes. For example, as previously highlighted, VPA is a well-recognised HDAC inhibitor, and pre-treatment with methionine has been shown to significantly reduce the incidence of spina bifida and other VPA-associated defects in mice [80].

4.4.6 Conclusion and future direction

In conclusion, the findings presented here offer a mechanistic explanation for the adverse neurodevelopmental effects caused by VPA. These effects are rooted in the dysregulation of gene transcription induced by the drug. The extent to which these transcriptional effects are linked to irreversible brain development or persistent changes encoded in the epigenetic makeup, or whether they are associated with a gradual restoration of normal brain transcription and function over time after birth, could be explored using the experimental approach described in this study. This underscores the critical importance of long-term follow-up studies on children born after gestational exposure to VPA.

Chapter 5

Discussion

5.1 Discussion

In the previous chapters, I have detailed three studies that utilised integrative genomic approaches to identify and evaluate the underlying biological mechanisms involved in cognitive disorders stemming from neurodegenerative and neurodevelopmental impairments. In this chapter, I will (a) provide a summary of the main findings from each study; (b) provide a summary of the primary mechanisms that were shared between these two types of cognitive disorders, as well as those that were unique to each; (c) I will address any biological or technical limitations that arose during the course of these analyses, while also discussing the notable advancements made in RNA-sequencing technologies; and (d) I will outline my research plans for future endeavours in this field.

5.1.1 Using transcriptomics profiling to study brain disorders: A summary of key findings from the current studies.

Transcriptomic analysis is a powerful tool for identifying gene expression signatures and associated biological processes in cognitive disorders Chapter 2, 3, 4. Although it cannot distinguish between causal and compensatory factors, it offers valuable insights into the molecular pathways underlying the pathogenesis of these disorders. This knowledge can aid in understanding the biological basis and help discover new therapeutic targets.

Below are summaries of the major findings, including novel discoveries, identified through transcriptomic profiling from the studies mentioned in this thesis:

Chapter 2: Single-nucleus transcriptomics identifies common and distinct molecular pathologies in Lewy body diseases.

1. In this study, transcriptional altercations were observed in multiple cell-types across several comparisons distinguishing the three LBDs under investigation. From differential expression analysis (Chapter 2.3.4; Figure 2.3), the following observations were made:

(a) there was a widespread dysregulation in neurons, as well as glial cell-types; (b) gene expression profiles of PDD and DLB were very similar (this likely stems from the fact that both conditions are cortical diseases and in contrast, PD without dementia primarily affects the motor system), while there were less transcriptional similarities between PD and PDD; (c) approximately around 36% of DEGs were found uniquely expressed in one comparison; (d) focusing on the disease-control comparison, it was evident that there was more down-regulation of genes in PDD and DLB and up-regulation of genes in PD (Figure 2.4), suggesting that PD may have a distinct transcriptional profile compared to the dementia groups.

2. Heritability enrichment analysis revealed a genetic association between DEGs in glial cells (up-regulated genes in OPCs and dysregulated (up and down) genes in astrocytes derived from PD versus control comparisons; $\text{FDR}_{\text{LDSC}} = 0.0076$ in OPC and $\text{FDR}_{\text{LDSC}} = 0.0085$ in astrocytes) and genetic determinants of PD age of onset (Chapter 2.3.7 and Figure 2.9). Although the involvement of OPCs has been shown in previous studies [39, 4], the role of astrocytes in PD age onset was also observed in the current study, Chapter 2.3.7.

3. A unique population of neurons associated with DLB was identified through clustering analysis (Chapter 2.3.7). These neurons were found to share a transcriptional profile similar to that of medium spiny neurons (MSN). As mentioned in the previous chapter, MSN are a type of GABAergic neurons primarily found in the striatum (a region involved in movement control). In PD, the progressive degeneration of dopaminergic neurons in the substantia nigra leads to a reduction in dopamine levels in the striatum, which in turn, contributes to the loss of MSN [342, 364]. Interestingly, these particular sub-types were not present in DLB groups (although extensively present in PDD group Figure 2.11), which mirrors earlier findings indicating that medium spiny neurons experience selective dendritic degeneration in DLB [354]. Furthermore, Skene et al. (2018) [283] were able to find a significant enrichment for schizophrenia in striatal medium spiny neurons using both LDSC and MAGMA. This may indicate that the loss of MSNs represents a convergent mechanism for the hallucinations seen in DLB and those in schizophrenia. However, what is unclear is the role of MSNs in PDD, since they are more abundantly present in the PDD

samples than PD samples. It is worth noting that this study represents the first instance in which these neurons have been identified and characterized at the transcriptional level (refer to Appendix A, Table 5 for full list of enriched functional terms).

Chapter 3: Uncovering Early-Stage Abnormal Tau Species-Related Gene Expression Changes Through Transcriptomic Profiling.

1. The study in Chapter 3 looked at transcriptional profiling of brain tissue samples from both tauPLA positive (tauPLA+; containing tau oligomers and NFTs) and tauPLA negative (tauPLA-) brain tissues from temporal cortex. Analysis of the gene expressions at the single cell-level revealed distinct and similar patterns between the two groups. Notably, there was a significant down-regulation of genes in excitatory neurons in tauPLA+ samples compared to tauPLA- samples. While it is widely recognized that the presence of NFTs (in samples with tauPLA+ and AT8+) leads to changes in gene expression patterns in neurons [97, 78], this study is unique in that it explores alterations in gene expression in the absence of NFTs and in the presence of tau multimers/oligomers (tauPLA+ and AT8- samples; 3.7).

2. The early activation of reactive astrocytes was observed in the study. It is well-established that reactive astrocytes play a crucial role in the progression of AD [286]. Nevertheless, the observation of their activation in the absence of NFTs has not been reported previously. To our knowledge, this is the first study to have shown similar transcription profile between astrocytes derived from samples that were both AT8+/tauPLA+, and those that were AT8-/tauPLA+, suggesting that reactive astrocytes are activated even in the absence of NFTs (refer to Appendix B, Figure 1 for average expression of marker genes across the three groups).

3. The findings of this study highlight the need for meticulous consideration when using control samples in transcriptomic investigations that involve aged or AD brains. Notably, this study demonstrates that the presence of newly discovered tau species (containing tau oligomers/multimers) can lead to significant disruptions in various pathways, even in the absence of NFTs. These results underscore the complexities of studying the transcriptome in AD and further emphasize the importance of rigorous experimental design to achieve a more nuanced understanding of the underlying molecular mechanisms of this devastating disease.

Chapter 4: Integrative genomics reveals pathogenic mediator of valproate-induced neurodevelopmental outcomes.

1. In this chapter, valproate exposure was shown to cause significant changes in gene expression in gestationally exposed pup brains, specifically down-regulating genes related to synaptic function and neurodevelopment. Furthermore, these differentially expressed genes were shown to be preferentially enriched for heritability to schizophrenia, bipolar disorder and IQ (Figure 4.6). This finding indicates that foetal valproate exposure can negatively affect the developing brain through drug-induced transcriptional dysregulation of genes that may be involved in neurodevelopment, synaptic dysfunction and protein homeostasis (Appendix C, Table 2 and 3 for full functional enrichment and LDSC result tables).

2. Through differential splicing analysis, some of the DEGs were shown to be also differentially spliced; this may suggest that the dysregulation of transcription induced by valproate is encoded epigenetically and may have enduring effects on human brain function (Chapter 4.4.4).

5.1.2 Convergent and divergent mechanisms underlying various types of neurodegenerative disorders.

Below is a summary of convergent and divergent mechanisms identified:

1. Although there were novel findings found in each chapter, most identified perturbed pathways in Chapter 2 (involving PD, PDD and DLB) and Chapter 3 (samples with NFTs (a feature of AD)) were similar. Examples of these pathways include axonal degeneration, mRNA splicing, synaptic organisation, autophagy, neuron death, phosphorylation, memory, mitochondrial function, and regulation of vesicle-mediated transport.

   Given the widely accepted notion that NFTs are a hallmark of dementias such as PDD and DLB [33, 13, 78], it is perhaps not surprising that the present study's findings support

previous observations made in similar studies [104, 131]. However, what is of particular interest is that the same molecular pathways were found to be perturbed in samples with few or no NFTs (Chapter 3.3.4 and Figure 3.7). This observation may indicate that tau aggregates (specifically tau oligomers/multimers) induced neurotoxicity represents an early molecular event in the pathogenesis of these diseases.

2. Another striking parallel between the two investigations was the observation of the down-regulation of genes (in the diseases tissue samples, namely PD, DLB, PDD and tauPLA+/AT8+ samples) in glial cell types, implying that not only are neurons susceptible to toxic tau species (in this case tau oligomers/multimers), but also glial cell-types. Specifically, several genes were found to be dysregulated in astrocytes (including reactive astrocyte marker genes) and oligodendrocytes within tauPLA+ brain tissues compared to controls (tauPLA-) (Figure 2.3 and Figure 3.5.). These findings suggest that further research into the involvement of glial cells in the aetiology of neurodegenerative diseases and their potential contribution to the spread of tau pathology is needed.

5.1.3 Overlapping and distinct perturbed pathways in neurodegeneration and neurodevelopment

Neurodevelopmental disorders (NDDs) and neurodegenerative diseases (NDs) are two distinct neurological conditions, with the former commonly manifesting during the critical period of brain development, and the latter typically presenting in later stages of life. As such, NDDs are often characterised by compromised brain development and function, while NDs are marked by the gradual degeneration of neurons (as seen in this study, for example, losses of medium spiny neurons-associated nuclei in DLB (Chapter 2.3.7) and related brain function).

1. Among the similarities between NDs and NDDs is the role of abnormal protein aggregation/synthesis. In NDs, misfolded proteins such as β-amyloid, tau, and α-synuclein accumulate in different regions of the brain, leading to neuronal dysfunction and selective neuronal death [57, 315, 125]. In NDDs such as ASD and schizophrenia, there is growing

evidence of altered synaptic plasticity [109, 30], and neuronal connectivity [109], which may be related to aberrant protein synthesis and degradation [189, 76]. These commonalities suggest that there may be shared mechanisms underlying protein homeostasis in both NDs and NDDs. These previous findings were mirrored in the current studies, in which terms such as protein-containing complex disassembly, protein secretion, regulation of protein polymerization were enriched for down-regulated genes across all disease groups (such as PD, PDD, DLB, tauPLA+/AT8+, VPA-Exposed groups) in comparison to control groups (2.5, Figure 2.6, Figure 3.7, Figure 4.5).

2. The involvement of GABAergic signalling in NDs and NDDs has been extensively studied in animal and human models [254, 199, 306, 156, 32]. It is widely recognized that GABAergic neurons play a crucial role in brain development and function. In NDDs, studies have shown that there may be alterations in GABAergic system [306, 126, 121] which leads to an imbalance between excitatory and inhibitory neurotransmission [306, 32]. This imbalance is believed to contribute to these disorders' cognitive and behavioural symptoms. Notably, the current study revealed that terms related to the regulation of GABAergic transmission were enriched for down-regulated genes in VPA-exposed pup brains.

   In contrast, it has been reported NDDs such as AD and PD exhibit a reduction in GABAergic interneurons in certain brain regions [347, 132, 231], in addition to the loss of other types of neurons. This loss of GABAergic neurons is known to contribute to the cognitive and motor symptoms observed in these disorders. Strikingly, in the current study, medium spiny neurons (a type of GABAergic neurons) were shown to be entirely lost in DLB (Figure 2.10).

3. Inflammatory responses and inflammation-associated pathways are common features of both NDs and NDDS. Studies have shown that chronic inflammation contributes to the pathophysiology of NDs and NDDs, with increased levels of pro-inflammatory cytokines observed in individuals with AD [294, 362] and ASD. Interestingly, however, the present study did not find any enrichment of inflammation-related pathways among the DEGs identified in VPA-exposed pup brains, although these pathways were enriched for DEGs

in PD, PDD and DLB.

4. In addition to inflammation, mitochondrial dysfunction is another hallmark of both NDs and NDDs [175, 276, 312]. However, the current study's perturbed pathways in VPA-exposed pup brains did not include mitochondrial dysfunction. While, several mitochondrial-associated pathways, including mitochondrion organization, apoptotic mitochondrial changes and mitochondrial transmembrane transport were enriched for DEGs from NDs (PD, PDD, DLB and tauPLA+/AT8+) versus control comparisons.

In summary, NDs and NDDs are two distinct categories of neurological conditions, each involving different underlying mechanisms and pathologies. In NDDs, pathways related to neurodevelopmental processes such as neuronal migration, differentiation, and neurogenesis are primarily affected, while in NDs, pathways related to neurodegenerative processes such as protein aggregation, autophagy, and neuroinflammation are mainly affected. However, despite their distinct differences, the current study has shown that NDDs and NDs share some commonalities in the perturbed pathways and processes that contribute to their respective pathologies. These similarities suggest that there may be some overlap in the underlying mechanisms of these disorders and highlight the need for further investigation into the shared pathways and potential therapeutic targets.

5.1.4 Challenges in transcriptomic data analysis arising from biological aspects

Ageing

Ageing is a biological process which results from the accumulation of diverse forms of molecular and cellular damage. The ageing brain undergoes a series of intricate changes, including a decline in cellular function, loss of synaptic connections, and a reduction in the production of neurotrophic factors [343, 124]. These age-related alterations contribute significantly to the onset and progression of NDs including AD and PD, by disrupting the delicate balance of the

brain's physiological and biochemical processes [124].

For example, one of the hallmark features of NDs as mentioned in previous chapter is the accumulation of misfolded protein in the brain, which can lead to the formation of aggregates and neuronal death [264, 323]. Age-related changes in protein turnover and clearance mechanisms can impair the ability of the brain to remove these toxic proteins, further exacerbating disease pathology [264]. Interestingly, in the current study, trajectory analysis (Chapter 3.3.6) demonstrated an increase expression pattern of reactive astrocyte marker genes (an astrocyte population which play a key role in clearing tau protein and other debris from the brain in response to injury or disease [180]) increased with age. This may be due to reactive astrocytes becoming dysfunctional and losing their ability to clear tau protein effectively as the brain ages. This may lead to the accumulation of tau protein and the formation of toxic tau species, such as NFTs, a hallmark of AD's pathology [141]. Notably, the current study found that these marker genes were activated even in the absence of NFTs (Appendix B, Figure 1), suggesting that tau oligomers may have a similar level of neurotoxicity as hyperphosphorylated NFTs.

In addition, age-related changes in the immune system can also contribute to neurodegeneration [124]. Microglia, specialised immune cells of the brain, play a crucial role in maintaining brain homeostasis and responding to injury and infection. However, with age, microglia can become dysfunctional, leading to chronic inflammation and further damage to neurons [194]. Unlike the astrocytes, early microglia activation was not observed in samples with only tau oligomers (and few to no NFTs), suggesting that astrocytes may have an early role in the pathogenesis of neurodegeneration than microglia.

Other age-related factors include: (i) sample heterogeneity: for example, the presence of senescent cells [161], which are cells that have stopped dividing due to a variety of stressors, such as damage to their DNA or other forms of cellular stress, can affect gene expression patterns and confound interpretation of results; (ii) comorbidities: due to the nature of ageing, individuals are more likely to develop multiple health conditions [73, 357], making it difficult to detangle the effects of different factors on disease development and progression; and (iii) other time-

dependent variables including, duration of disease, time since intervention, length of exposure to treatment and survivorship bias (older individuals may be healthier or have a different health trajectory)[9].

Separating the effects of ageing from the effects of the disease in transcriptomic based studies.

To address the challenges related to age-related factors in case-control transcriptomic studies, careful study design is essential. Important considerations for designing such studies include the use of appropriate statistical methods to account for potential confounding variables, and the utilisation of reference datasets containing age-matched controls to ensure that age-related effects are accurately accounted for.

Stratifying samples by age groups can be an apparent choice for case-control studies. However, this approach may lead to smaller sample sizes within each age group, which can lower the statistical power and make it challenging to detect significant differences in gene expression between groups. Apart from stratifying samples into age groups, researchers have applied several means to account for age-related changes in gene expression. For example:

1. Age can be incorporated as a covariate in statistical models to control its effects on gene expression (as done in the current study). This approach allows the identification of disease-related changes in gene expression that are independent of age. However, this may not completely eliminate the confounding effect of age-related changes on gene expression, as there may be other variables that are also affected by aging and contribute to gene expression changes.

2. Recently, machine learning algorithms have been utilised to identify age-related patterns in gene expression data [163, 85]. Briefly, the algorithms are trained on gene expression profiles from individuals of different ages, and then the model created is used to predict gene expression changes associated with ageing. Hence, this could be one way for accounting the effect of ageing on gene expression.

In addition to age, PMI can pose a challenge in transcriptomic studies, as variations in PMI among samples can impact gene expression patterns. Recently, Zhu et al., 2017 [366] conducted a systematic analysis of gene expression patterns associated with PMI in human tissues and revealed tissue and cell-type specific mRNA degradation profiles. Although PMI is currently used as a covariate in statistical methods, identifying PMI-associated genes could enhance the interpretation of gene expression patterns linked to disease status.

In both studies related to NDs and NDDs, sex-bias may possess a challenge when it comes to identifying sex-based changes or disease-driven expression changes. This is because the expression of genes can vary between males and females due to differences in hormonal and genetic factors [314, 260].

For instance, in AD, women are known to have a higher prevalence and faster progression than men. One possibility could be due to the protective effect of testosterone in males as demonstrated by Rosario et al., 2011 [260] in which the authors found inverse relationship between brain levels of testosterone and soluble Aβ. In contrast, as seen in the current study, in DLB men were more likely to develop the disease than women. However, there are conflicting reports have been seen in other studies [95]. Therefore, failing to account for sex differences in gene expression patterns could lead to a lack of understanding of the underlying disease mechanisms in both males and females.

Similarly, in ASD, males are known to be affected more frequently and severely than females [95]. This suggests that there could be sex-specific mechanisms underlying the disorder that are not accounted for when sex differences are not taken into consideration. Therefore, to overcome these challenges, it is important to design studies that include both males and females in sufficient numbers and to statistically account for sex differences in gene expression. This can be done by stratifying (having equal/equivalent male and female samples) or including sex as a covariate in the statistical models, as done in this study.

5.1.5 Technical differences between single-cell and single-nucleus RNA-Seq

The development of scRNA-Seq technology has revolutionised our understanding of disease by revealing previously undiscovered cellular heterogeneity and novel cell-types through the analysis of gene expression patterns at the single-cell level [127, 181]. This capability enables researchers to examine gene expression across a vast range of cell subpopulations, revealing unique cell-types that may play pivotal roles in disease pathogenesis and progression [285]. Additionally, scRNA-Seq can identify key regulatory genes and pathways that are dysregulated in disease, uncovering potential molecular mechanisms that lead to disease pathology or those involved in the progression of the disease. Hence providing new insights into disease pathogenesis and potential targets for therapeutic intervention.

Although, scRNA-Sequencing gives a comprehensive understanding of cellular heterogeneity which allows the identification of individual cell-types, splicing events and different isoforms, which may be missed in bulk RNA-Sequencing, scRNA-Seq requires the isolation of viable single cells, which can be technically challenging and may introduce bias due to the isolation process [160]. In order to overcome this technical challenge, snRNA-Seq was developed as an alternative to scRNA-Seq for analysing gene expression in samples with degraded RNA (or low RNA content) [160]. By using the nuclei of the cells rather than the entire cell, snRNA-Seq enables the analysis of gene expression in intact tissue samples at a single-cell resolution. One of the main advantages of snRNA-Seq is that it can be performed on frozen, fixed, or even archived tissues, which may be the only available samples for certain studies [160, 167].

5.1.6 The limitations of single-nucleus RNA-Sequencing

Over the last few years, numerous studies have used snRNA-Seq to unravel the complexities of the brain. These studies have focused on understanding the transcriptional and epigenetic states of the human brain, identifying neuronal subtypes, and or even exploring the single-

cell transcriptional landscape of mammalian organogenesis. Notably, Habib et al., 2020 [111] identified rare cell types in the mouse brain using snRNA-Seq, while Lake et al., 2016 [166] employed the same method to uncover novel neuronal subtypes using nuclei from post-mortem brain. In another study, Zeng et al., 2016 [359] conducted both snRNA-Seq and scRNA-Seq on immortalized human myoblasts and found them to be comparable, with a distinct enrichment for long non-coding RNAs in snRNA-seq.

However, it is worth noting that snRNA-Seq has several limitations compared to scRNA-Seq including:

1. The potential loss of information due to the isolation of nuclei instead of whole cells (low sensitivity); thus, leading to underestimation or loss of expression patterns from rare cell-types.

2. Low coverage, due to low capture efficiency of nuclei and higher dropout rates, could reduce the ability to detect lowly expressed genes, or low-abundance transcripts, which could lead to inaccurate quantification of gene expression levels.

3. Damage to cells due to the isolation process which could lead to unreliable or lower quality data [44].

4. Batch effects due to the use of different isolation protocols and library preparation methods making it challenging to compare results across multiple studies.

5. Challenges in differentiating between transcripts present in the nucleus and cytoplasm [353].

6. High level of background noise (specifically in 10x-based platforms) due to the presence of mitochondrial RNA (mtRNA) and ribosomal RNA (rRNA) in the nuclei [44]. mtRNA and rRNA are highly abundant in the nucleus, and their transcripts are captured along with the mRNA during the snRNA-Seq process, leading to a large proportion of reads being derived from these RNAs [193, 369]. This can result in decreased sensitivity and accuracy of downstream analyses, as well as an increased cost due to the need for deeper

sequencing to achieve sufficient coverage. For the current study, `CellBender` was applied to overcome this challenge. In conclusion, snRNA-Seq remains a valuable tool for studying gene expression in individual nuclei and has led to many significant discoveries in brain sciences. However, researchers should carefully design their studies and when choosing which technology to use. Furthermore, it is important to take into account the appropriate normalisation and statistical methods when analysing the data to avoid common pitfalls in differential expression analysis.

5.1.7 Other technical limitations

(i) Proteomics

RNA-seq experiments are susceptible to technical biases, such as library preparation artefacts and amplification biases, which can lead to false positives and distort the representation of specific RNA species [280]. One notable limitation of the current studies is their exclusive reliance on mRNA expression levels, thereby overlooking the assessment of protein levels. Until recently, it was widely assumed in systems biology literature that mRNA and protein expressions measured from a tissue exhibited a proportional relationship. However, multiple studies analysing mRNA and protein expression data from the same cells under similar conditions have challenged this assumption, revealing a lack of high correlation between these two distinct molecular entities with unique properties and behaviour within cells [50, 114, 239, 108].

Several factors contribute to the disparities between mRNA and protein measurements. For example, various post-transcriptional regulations, including mRNA stability and protein degradation, can influence the abundance of proteins without corresponding changes in mRNA levels [350, 114, 326]. In addition, protein-protein interactions, post-translational modifications, such as phosphorylation, glycosylation, and acetylation, can impact the function and localisation of proteins, however, these modifications are not directly captured by mRNA measurements [11, 138, 68]. In a recent study, Takemon et al.(2021) [303] conducted a study in which they measured mRNA and protein levels in mice across various ages. The findings revealed that

age-related alterations in protein levels were not solely attributed to corresponding changes in mRNA levels.

In general, the analysis of mRNA transcripts is considerably more accessible compared to proteins. Transcripts can be readily detected and characterised using techniques such as northern blotting, PCR amplification, and cDNA sequencing. Whereas, studying proteins requires a range of costly and labour-intensive methodologies, including mass spectrometry [311]. Therefore, despite the limitations, transcriptomics continues to be a valuable tool in unravelling biological processes, gene regulation, and disease mechanisms. Integrating protein measurements and other omics data can further enhance our understanding of complex cellular processes and facilitate more comprehensive investigations in the future.

(ii) Differential expression analysis methods

New tools for analysing gene expression, along with advances in scRNA-Seq technology, are greatly improving our ability to study gene expression patterns with more accuracy than ever before. While scRNA-Seq/snRNA-Seq enables the measurement of gene expression at the single-cell level, DEA allows for the identification of genes that are differentially expressed between two or more conditions, cell-types or cell-states, providing insight into the molecular basis of disease pathogenesis and progression.

However, current DEA methods have several limitations including:

1. The assumption (especially in MAST) that the genes are independently and identically distributed across cells, which has been shown to not be the case in some studies [112].Furthermore, MAST assumes genes across cells follow a negative binomial distribution, which may not always be the case. This can result in inflated test statistics or inaccurate estimation of expression variability and increased number of false positives in DEA [112, 333, 291].

2. To address these issues, alternative methods have been developed that do not make assumptions about the distribution of gene expression levels, such as the non-parametric

approach scVI [188], which uses deep learning to model the gene expression distribution. However, machine-learning based methods also has limitations, including difficulty in identifying rare cell-types due to their absence in the training data; and requiring significant computational resources to train and implement, which may limit its accessibility for some researchers.

5.1.8 Alternatives to using single-nucleus RNA-Sequencing based differential expression analysis methods:

Pseudobulk methods are becoming widely used for differential expression analysis using sn/scRNA-Seq data [220]. These methods involve averaging the expression values of cells/nuclei within the same individual, such as cells from the same tissue or cell type, to obtain a bulk-like expression profile. However, these methods have several limitations. For example, they may fail to detect subtle gene expression changes and cell-to-cell variation within a given sample [367], potentially leading to the loss of important information regarding subpopulation-specific gene expression changes and rare cell types with distinct transcriptional profiles, as seen in the current study.

Another limitation of pseudobulk methods is the assumption that the same cell-type composition/proportion is the same across samples, which may not be the case in reality. This can mask differences in cell-type composition between samples and confound identification of DEGs. Additionally, pseudobulk methods may not perform well in detecting differential expression in lowly expressed genes.

Despite these limitations, pseudobulk methods can still be a useful tool for DEA. However, researchers should carefully consider the limitations when deciding to use the methods and supplement them with other approaches, such as MAST or other sn/scRNA-Seq based methods, to obtain a more comprehensive understanding of gene expression changes.

5.1.9 Future work

Most of the analyses carried out in the current study rely heavily on identifying potentially disease-related genes by comparing their expression levels between groups of individuals with or without the pathology. If reverse causality is present, the observed difference in gene expression could potentially be due to the disease-causing changes in gene expression rather than the gene expression changes causing the disease. Reverse causality is a type of bias that can occur in research (including ones where DEA is used) when the cause-and-effect relationship between two variables is incorrectly assumed due to an observational association. For example, instead of A (gene expression) causing B (disease), it may be that B causes A [226].

In order to explore potential causal relationships between variables and mitigate the impact confounding factors such as reverse causality, it is important to complement DEA with other additional methodologies such as expression quantitative trait loci (eQTLs) analysis, Mendelian Randomisation (MR), colocalization and mediation analysis [226]. By utilising these methods, researchers can better assess whether a given exposure (such as drug exposure) or trait is causally linked to a particular outcome of interest, such as disease risk or progression.

A brief explanation of these analyses is as follows:

1. expression quantitative trait loci (eQTLs) analysis is a widely used method which was first introduced as a means of identifying relationships between genetic polymorphisms and gene expression variation. These eQTLs (genetic variants associated with gene expression levels) are used in colocalization analysis (CA) to investigate the potential causal relationships between genetic variants and complex traits, including diseases [226, 246].

2. Colocalisation analysis (CA) is a method used to investigate whether two traits or phenotypes are associated with the same causal genetic variant; and to test if the same genetic variants are found to be associated with both the disease and the gene expression, this suggests that the genetic variant is likely to be causal for the disease through its effect on the gene expression [339].

3. MR is used to test the causal relationship between the genetic variants disease by utilising the variant as an instrumental variable. Briefly MR is a statistical method that uses genetic variants to determine the causal relationship between a risk factor and an outcome of interest (for example, a disease). MR utilises the random distribution of genetic variants during meiosis (gamete formation), leading to their independence from confounding factors such as lifestyle or environmental factors [173]. These genetic variants are used as instrumental variables to estimate the causal effect of the risk factor on the outcome.

4. Mediation analysis is a method used to investigate the mechanism or pathway through which an independent variable influences a dependent variable. The mediator variable is a variable that explains or accounts for the relationship between the independent and dependent variables. For example, when studying genes expression change, age and disease outcome; gene expression level is a mediator variable, as changes in gene expression levels with age could affect disease risk; age is the independent variable (or predictor variable) as it could predict disease outcome; and finally, disease outcome can be the dependent variable (or outcome variable) as it is the variable of interest and could be predicted by age and gene expression levels. Other variables such as sex, smoking, diet, exercise and race could be included as covariates in the analysis to control for potential confounding effects [358].

In summary, integrating these methods in a broad-ranging causal inference framework, as I plan to do in my future analyses, can provide more reliable and accurate assessment into the underlying biological mechanisms and the potential for therapeutic interventions.

Appendix

Please refer to `https://github.com/rahfel/Appendices` for the following xls. and png files:

1. Appendix A, Table 1.Sample demographics and clinical and pathological measures
2. Appendix A, Table 2.Quality control metrics
3. Appendix A, Table 3.The full number of nuclei per subject
4. Appendix A, Table 4.Cell-type specific differential expression analyses results
5. Appendix A, Table 5.Functional pathway enrichments across all comparisons
6. Appendix A, Table 6.Functional pathway enrichments across all comparisons (PD-linked pathways)
7. Appendix A, Table 7.HMAGMA full results table
8. Appendix A, Table 8.sLDSC full results table
9. Appendix B, Table 1.Clinical and pathological features of the cohort
10. Appendix B, Table 2.Quality measures for read mapping and sequencing
11. Appendix B, Table 3.List of nuclei derived per subject
12. Appendix B, Table 4.Cell-type specific differential expression analyses results
13. Appendix B, Table 5.Functional pathway enrichments across all comparisons
14. Appendix B, Table 6.LDSC full results table
15. Appendix B, Table 7.HMAGMA full results table
16. Appendix B, Table 8.Number of genes per astrocyte module
17. Appendix B, Table 9.Enrichment of biological processes pathways result for astrocyte modules

18. Appendix B, Table 10.Fisher's exact test result for enrichment of reactive astrocyte marker genes

19. Appendix B, Table 11.Enrichment of heritabilty result table for astrocyte modules

20. Appendix B, Table 12.Number of genes per excitatory module

21. Appendix B, Table 13.Enrichment of biological processes pathways result for excitatory modules

22. Appendix B, Table 14.Enrichment of heritabilty result table for excitatory modules

23. Appendix B, Figure 1.Reactive astrocyte average gene expression per group

24. Appendix B, Figure 2.Overlap of DEGs and neurodegenerative associated genes from Genomics England gene panel.

25. Appendix C, Table 1.Differential expression analyses results table

26. Appendix C, Table 2.Functional enrichment.Overrepresentation analysis results table

27. Appendix C, Table 3.LDSC full results table

28. Appendix C, Table 4.DWLS reuslt table

29. Appendix C, Table 5.Differential splicing analyses result table

Bibliography

[1] Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. *Nature communications 9*, 1 (2018), 5269. 4.2.5, 4.3.5

[2] Aarsland, D., and Kurz, M. W. The epidemiology of dementia associated with parkinson disease. *Journal of the Neurological Sciences 289* (2 2010), 18–22. 2.1

[3] Adams, J. N., Maass, A., Harrison, T. M., Baker, S. L., and Jagust, W. J. Cortical tau deposition follows patterns of entorhinal functional connectivity in aging. *eLife 8* (9 2019). 3.1.3

[4] Agarwal, D., Sandor, C., Volpato, V., Caffrey, T. M., Monzón-Sandoval, J., Bowden, R., Alegre-Abarrategui, J., Wade-Martins, R., and Webber, C. A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nature Communications 11* (8 2020), 4183. 2.4.4, 2

[5] Akbarian, S., Liu, C., and Knowles, J. The psychencode project. *Nat Neurosci 18*, 12, 1707–1712. 11. 1.1

[6] Amer-Sarsour, F., Kordonsky, A., Berdichevsky, Y., Prag, G., and Ashkenazi, A. Deubiquitylating enzymes in neuronal health and disease. *Cell Death Dis 12*, 1. 213. 2

[7] Anang, J. B. M., Gagnon, J.-F., Bertrand, J.-A., Romenets, S. R., Latreille, V., Panisset, M., Montplaisir, J., and Postuma, R. B. Predictors

of dementia in parkinson disease: a prospective cohort study. *Neurology 83* (9 2014), 1253–60. 2.1

[8] Anders, S., and Huber, W. Differential expression analysis for sequence count data. *Genome Biology 11* (10 2010), R106. 1.1.3

[9] Anderson, C. D., Nalls, M. A., Biffi, A., Rost, N. S., Greenberg, S. M., Singleton, A. B., Meschia, J. F., and Rosand, J. The effect of survival bias on case-control genetic association studies of highly lethal diseases. *Circulation: Cardiovascular Genetics 4* (4 2011), 188–196. 5.1.4

[10] Andrews, S. Fastqc: a quality control tool for high throughput sequence data. 1.1.3, 4.2.2

[11] Archer, T. C., Ehrenberger, T., Mundt, F., Gold, M. P., Krug, K., Mah, C. K., Mahoney, E. L., Daniel, C. J., LeNail, A., Ramamoorthy, D., Mertins, P., Mani, D., Zhang, H., Gillette, M. A., Clauser, K., Noble, M., Tang, L. C., Pierre-François, J., Silterra, J., Jensen, J., Tamayo, P., Korshunov, A., Pfister, S. M., Kool, M., Northcott, P. A., Sears, R. C., Lipton, J. O., Carr, S. A., Mesirov, J. P., Pomeroy, S. L., and Fraenkel, E. Proteomics, post-translational modifications, and integrative analyses reveal molecular heterogeneity within medulloblastoma subgroups. *Cancer Cell 34* (9 2018), 396–410.e8. 5.1.7

[12] Armstrong, M. J. Advances in dementia with lewy bodies. *Therapeutic advances in neurological disorders 14* (2021), 17562864211057666. 2.1

[13] Arriagada, P. V., Growdon, J. H., Hedley-Whyte, E. T., and Hyman, B. T. Neurofibrillary tangles but not senile plaques parallel duration and severity of alzheimer's disease. *Neurology 42* (3 1992), 631–631. 3.1.2, 1

[14] Avery, L. B., and Bumpus, N. N. Valproic acid is a novel activator of amp-activated protein kinase and decreases liver mass, hepatic fat accumulation, and serum glucose in obese mice. *Molecular Pharmacology 85* (1 2014). 4.3.4

[15] Ayers, J. I., Giasson, B. I., and Borchelt, D. R. Prion-like spreading in tauopathies. *Biological Psychiatry 83* (2 2018), 337–346. 3.1.2, 3.1.3

[16] Bandres-Ciga, S., Saez-Atienzar, S., and Kim, J. Large-scale pathway specific polygenic risk and transcriptomic community network analysis identifies novel functional pathways in parkinson disease. *Acta Neuropathol 140*, 3, 341–358. 160. 2.3.5, 2.3.6

[17] Barbier, P., Zejneli, O., Martinho, M., Lasorsa, A., Belle, V., Smet-Nocca, C., Tsvetkov, P. O., Devred, F., and Landrieu, I. Role of tau as a microtubule-associated protein: Structural and functional aspects. *Frontiers in Aging Neuroscience 11* (8 2019). 3.1.2

[18] Bardoni, B., Abekhoukh, S., Zongaro, S., and Melko, M. Intellectual disabilities, neuronal posttranscriptional rna metabolism, and rna-binding proteins. *In 2012*, 29–51. 60. 1.2.1

[19] Barkas, N., Petukhov, V., and Nikolaeva, D. Joint analysis of heterogeneous single-cell rna-seq dataset collections. *Nat Methods 16*, 8, 695–698. 147. 1.1.3, 2.2.5

[20] Becht, E., McInnes, L., and Healy, J. Dimensionality reduction for visualizing single-cell data using umap. *Nat Biotechnol 37*, 1, 38–44. 31. 1.1.3

[21] Bengoa-Vergniory, N., Velentza-Almpani, E., Silva, A. M., Scott, C., Vargas-Caballero, M., Sastre, M., Wade-Martins, R., and Alegre-Abarrategui, J. Tau-proximity ligation assay reveals extensive previously undetected pathology prior to neurofibrillary tangles in preclinical alzheimer's disease. *Acta Neuropathologica Communications 9*, 1 (2021), 1–20. 3.1.5

[22] Berendsen, S., Frijlink, E., Kroonen, J., Spliet, W. G. M., van Hecke, W., Seute, T., Snijders, T. J., and Robe, P. A. Effects of valproic acid on histone deacetylase inhibition in vitro and in glioblastoma patient samples. *Neuro-Oncology Advances 1* (5 2019). 4.1.2, 4.3.4

[23] Björklund, O., Kahlström, J., Salmi, P., and Fredholm, B. Perinatal caffeine, acting on maternal adenosine a1 receptors, causes long-lasting behavioral changes in mouse offspring. *PLoS One 3*, 12. 231. 4.1.2

[24] Blauwendraat, C., Heilbron, K., Vallerga, C. L., Bandres-Ciga, S., von Coelln, R., Pihlstrøm, L., Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Siitonen, A., Iwaki, H., Leonard, H., Noyce, A. J., Tan, M., Gibbs, J. R., Hernandez, D. G., Scholz, S. W., Jankovic, J., Shulman, L. M., Lesage, S., Corvol, J., Brice, A., van Hilten, J. J., Marinus, J., Eerola-Rautio, J., Tienari, P., Majamaa, K., Toft, M., Grosset, D. G., Gasser, T., Heutink, P., Shulman, J. M., Wood, N., Hardy, J., Morris, H. R., Hinds, D. A., Gratten, J., Visscher, P. M., Gan-Or, Z., Nalls, M. A., and Singleton, A. B. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and -synuclein mechanisms. *Movement Disorders 34* (6 2019), 866–875. 2.1.2, 2.2.9, 2.3.7

[25] Blauwendraat, C., Nalls, M. A., and Singleton, A. B. The genetic architecture of parkinson's disease. *The Lancet Neurology 19* (2 2020), 170–178. 2.1.2, 2.3.5

[26] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment 2008*, 10 (2008), P10008. 1.1.3

[27] Bohnen, N. I., Müller, M. L. T. M., and Frey, K. A. Molecular imaging and updated diagnostic criteria in lewy body dementias. *Current neurology and neuroscience reports 17* (8 2017), 73. 2.1.1

[28] Bondi, M. W., Edmonds, E. C., and Salmon, D. P. Alzheimer's disease: Past, present, and future. *Journal of the International Neuropsychological Society 23* (10 2017), 818–831. 3.1.1

[29] Boukhris, T., Sheehy, O., Mottron, L., and Bérard, A. Antidepressant use during pregnancy and the risk of autism spectrum disorder in children. *JAMA Pediatrics 170* (2 2016). 4.4.5

[30] Bourgeron, T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nature Reviews Neuroscience 16* (9 2015), 551–563. 1

[31] Bourin, M. Mechanism of action of valproic acid and its derivatives. *SOJ Pharmacy Pharmaceutical Sciences* (2020). 4.1.2

[32] Bouzamondo-Bernstein, E., Hopkins, S. D., Spilman, P., Uyehara-Lock, J., Deering, C., Safar, J., Prusiner, S. B., Ralston III, H. J., and Dearmond, S. J. The neurodegeneration sequence in prion diseases: evidence from functional, morphological and ultrastructural studies of the gabaergic system. *Journal of Neuropathology & Experimental Neurology 63*, 8 (2004), 882–899. 2

[33] Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., and Tredici, K. D. Staging of alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathologica 112* (10 2006), 389–404. 3.1.4, 1

[34] Braak, H., and Tredici, K. D. Neuropathological staging of brain pathology in sporadic parkinson's disease: Separating the wheat from the chaff. *Journal of Parkinson's disease 7* (2017), S71–S85. 2.1.1

[35] Braak, H., Tredici, K. D., Rüb, U., de Vos, R. A., Steur, E. N. J., and Braak, E. Staging of brain pathology related to sporadic parkinson's disease. *Neurobiology of Aging 24* (3 2003), 197–211. 2.1.1

[36] Brown, R., Lockwood, A., and Sonawane, B. Neurodegenerative diseases: An overview of environmental risk factors. *Environ Health Perspect 113*, 9, 1250–1256. 1.2.2

[37] Brunello, C. A., Merezhko, M., Uronen, R.-L., and Huttunen, H. J. Mechanisms of secretion and spreading of pathological tau protein. *Cellular and Molecular Life Sciences 77* (5 2020), 1721–1744. 3.1.2

[38] Bruno, V., Sortino, M. A., Scapagnini, U., Nicoletti, F., and Canonico, P. L. Antidegenerative effects of mg(2+)-valproate in cultured cerebellar neurons. *Functional neurology 10* (1995), 121–30. 4.1.2

[39] Bryois, J., Skene, N. G., Hansen, T. F., Kogelman, L. J. A., Watson, H. J., Liu, Z., Adan, R., Alfredsson, L., Ando, T., Andreassen, O., Baker, J., Bergen, A., Berrettini, W., Birgegård, A., Boden, J., Boehm, I., Boni, C., Perica, V. B., Brandt, H., Breen, G., Bryois, J., Buehren, K., Bulik, C., Burghardt, R., Cassina, M., Cichon, S., Clementi, M., Coleman, J., Cone, R., Courtet, P., Crawford, S., Crow, S., Crowley, J., Danner, U., Davis, O., de Zwaan, M., Dedoussis, G., Degortes, D., DeSocio, J., Dick, D., Dikeos, D., Dina, C., Dmitrzak-Weglarz, M., Martinez, E. D., Duncan, L., Egberts, K., Ehrlich, S., Escaramís, G., Esko, T., Estivill, X., Farmer, A., Favaro, A., Fernández-Aranda, F., Fichter, M., Fischer, K., Föcker, M., Foretova, L., Forstner, A., Forzan, M., Franklin, C., Gallinger, S., Gaspar, H., Giegling, I., Giuranna, J., Giusti-Rodríquez, P., Gonidakis, F., Gordon, S., Gorwood, P., Mayora, M. G., Grove, J., Guillaume, S., Guo, Y., Hakonarson, H., Halmi, K., Hanscombe, K., Hatzikotoulas, K., Hauser, J., Hebebrand, J., Helder, S., Henders, A., Herms, S., Herpertz-Dahlmann, B., Herzog, W., Hinney, A., Horwood, L. J., Hübel, C., Huckins, L., Hudson, J., Imgart, H., Inoko, H., Janout, V., Jiménez-Murcia, S., Johnson, C., Jordan, J., Julià, A., Juréus, A., Kalsi, G., Kaminská, D., Kaplan, A., Kaprio, J., Karhunen, L., Karwautz, A., Kas, M., Kaye, W., Kennedy, J., Kennedy, M., Keski-Rahkonen, A., Kiezebrink, K., Kim, Y.-R., Kirk, K., Klareskog, L., Klump, K., Knudsen, G. P., Via, M. L., Landén, M., Larsen, J., Hellard, S. L., Leppä, V., Levitan, R., Li, D., Lichtenstein, P., Lilenfeld, L., Lin, B. D., Lissowska, J., Luykx, J., Magistretti, P., Maj, M., Mannik, K., Marsal, S., Marshall, C., Martin, N., Mattheisen, M., Mattingsdal, M., McDevitt, S., McGuffin, P., Medland, S., Metspalu, A., Meulenbelt, I., Micali, N., Mitchell, J., Mitchell, K.,

Monteleone, P., Monteleone, A. M., Montgomery, G., Mortensen, P. B., Munn-Chernoff, M., Nacmias, B., Navratilova, M., Norring, C., Ntalla, I., Olsen, C., Ophoff, R., O'Toole, J., Padyukov, L., Palotie, A., Pantel, J., Papezova, H., Parker, R., Pearson, J., Pedersen, N., Petersen, L., Pinto, D., Purves, K., Rabionet, R., Raevuori, A., Ramoz, N., Reichborn-Kjennerud, T., Ricca, V., Ripatti, S., Ripke, S., Ritschel, F., Roberts, M., Rotondo, A., Rujescu, D., Rybakowski, F., Santonastaso, P., Scherag, A., Scherer, S., Schmidt, U., Schork, N., Schosser, A., Seitz, J., Slachtova, L., Slagboom, P. E., 't Landt, M. S.-O., Slopien, A., Sorbi, S., Strober, M., Stuber, G., Sullivan, P., Świątkowska, B., Szatkiewicz, J., Tachmazidou, I., Tenconi, E., Thornton, L., Tortorella, A., Tozzi, F., Treasure, J., Tsitsika, A., Tyszkiewicz-Nwafor, M., Tziouvas, K., van Elburg, A., van Furth, E., Wade, T., Wagner, G., Walton, E., Watson, H., Werge, T., Whiteman, D., Widen, E., Woodside, D. B., Yao, S., Yilmaz, Z., Zeggini, E., Zerwas, S., Zipfel, S., Anttila, V., Artto, V., Belin, A. C., de Boer, I., Boomsma, D. I., Børte, S., Chasman, D. I., Cherkas, L., Christensen, A. F., Cormand, B., Cuenca-Leon, E., Davey-Smith, G., Dichgans, M., van Duijn, C., Esko, T., Esserlind, A. L., Ferrari, M., Frants, R. R., Freilinger, T., Furlotte, N., Gormley, P., Griffiths, L., Hamalainen, E., Hansen, T. F., Hiekkala, M., Ikram, M. A., Ingason, A., Järvelin, M.-R., Kajanne, R., Kallela, M., Kaprio, J., Kaunisto, M., Kogelman, L. J. A., Kubisch, C., Kurki, M., Kurth, T., Launer, L., Lehtimaki, T., Lessel, D., Ligthart, L., Litterman, N., van den Maagdenberg, A., Macaya, A., Malik, R., Mangino, M., McMahon, G., Muller-Myhsok, B., Neale, B. M., Northover, C., Nyholt, D. R., Olesen, J., Palotie, A., Palta, P., Pedersen, L., Pedersen, N., Posthuma, D., Pozo-Rosich, P., Pressman, A., Raitakari, O., Schürks, M., Sintas, C., Stefansson, K., Stefansson, H., Steinberg, S., Strachan, D., Terwindt, G., Vila-Pueyo, M., Wessman, M., Winsvold, B. S., Zhao, H., Zwart, J. A., Agee, M., Alipanahi, B., Auton, A., Bell, R., Bryc, K.,

Elson, S., Fontanillas, P., Furlotte, N., Heilbron, K., Hinds, D., Huber, K., Kleinman, A., Litterman, N., McCreight, J., McIntyre, M., Mountain, J., Noblin, E., Northover, C., Pitts, S., Sathirapongsasuti, J., Sazonova, O., Shelton, J., Shringarpure, S., Tian, C., Tung, J., Vacic, V., Wilson, C., Brueggeman, L., Breen, G., Bulik, C. M., Arenas, E., Hjerling-Leffler, J., and Sullivan, P. F. Genetic identification of cell types underlying brain complex traits yields insights into the etiology of parkinson's disease. *Nature Genetics 52* (5 2020), 482–493. 2.4.4, 2

[40] Bulik-Sullivan, B., Loh, P., and Finucane, H. Ld score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet 47*, 3, 291–295. 47. 1.1.3, 3.2.11, 4.3.5

[41] Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology 36* (5 2018), 411–420. 1.1.3, 2.2.3, 3.2.7

[42] Butterfield, D. A., and Halliwell, B. Oxidative stress, dysfunctional glucose metabolism and alzheimer disease. *Nature Reviews Neuroscience 20* (3 2019), 148–160. 3.1.2

[43] CA, L., JM, M., T, H., and MAGMA, P. D. Generalized gene-set analysis of gwas data. *PLoS Comput Biol 11*, 4. 46. 1.1.3

[44] Caglayan, E., Liu, Y., and Konopka, G. Neuronal ambient rna contamination causes misinterpreted and masked cell types in brain single-nuclei datasets. *Neuron 110* (12 2022), 4043–4056.e5. 3, 6

[45] Cannoodt, R., Saelens, W., and Saeys, Y. Computational methods for trajectory inference from single-cell transcriptomics. *Eur J Immunol 46*, 11, 2496–2506. 49. 1.1.3

[46] Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D. M., Hill, A. J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F. J., Trapnell, C., and

Shendure, J. The single-cell transcriptional landscape of mammalian organogenesis. *Nature 566* (2 2019), 496–502. 1.1.3, 3.2.12, 3.3.6

[47] Caponigro, F., Gennaro, E. D., Ionna, F., Longo, F., Aversa, C., Pavone, E., Maglione, M. G., Marzo, M. D., Muto, P., Cavalcanti, E., Petrillo, A., Sandomenico, F., Maiolino, P., D'Aniello, R., Botti, G., Cecio, R. D., Losito, N. S., Scala, S., Trotta, A., Zotti, A. I., Bruzzese, F., Daponte, A., Calogero, E., Montano, M., Pontone, M., Feo, G. D., Perri, F., and Budillon, A. Phase ii clinical study of valproic acid plus cisplatin and cetuximab in recurrent and/or metastatic squamous cell carcinoma of head and neck-v-chance trial. *BMC Cancer 16* (12 2016), 918. 4.1.2

[48] Casillas-Espinosa, P., Powell, K., and Zhu, M. Evaluating whole genome sequence data from the genetic absence epilepsy rat from strasbourg and its related non-epileptic strain. *PLoS One 12*, 7. 271. 4.2.1, 4.3.1

[49] Chau, D. K.-F., Choi, A. Y.-T., Yang, W., Leung, W. N., and Chan, C. W. Downregulation of glutamatergic and gabaergic proteins in valproric acid associated social impairment during adolescence in mice. *Behavioural Brain Research 316* (1 2017). 4.1.2, 4.1.3

[50] Chen, G., Gharib, T. G., Huang, C.-C., Taylor, J. M., Misek, D. E., Kardia, S. L., Giordano, T. J., Iannettoni, M. D., Orringer, M. B., Hanash, S. M., and Beer, D. G. Discordant protein and mrna expression in lung adenocarcinomas. *Molecular Cellular Proteomics 1* (4 2002), 304–313. 5.1.7

[51] Chen, Y., Lu, J., Pan, H., Zhang, Y., Wu, H., Xu, K., Liu, X., Jiang, Y., Bao, X., Yao, Z., Ding, K., Lo, W. H. Y., Qiang, B., Chan, P., Shen, Y., and Wu, X. Association between genetic variation ofcacna1h and childhood absence epilepsy. *Annals of Neurology 54* (8 2003), 239–243. 1

[52] Chia, R., Sabir, M. S., Bandres-Ciga, S., Saez-Atienzar, S., Reynolds, R. H., Gustavsson, E., Walton, R. L., Ahmed, S., Viollet, C., Ding, J., Makari-

ous, M. B., Diez-Fairen, M., Portley, M. K., Shah, Z., Abramzon, Y., Hernandez, D. G., Blauwendraat, C., Stone, D. J., Eicher, J., Parkkinen, L., Ansorge, O., Clark, L., Honig, L. S., Marder, K., Lemstra, A., George-Hyslop, P. S., Londos, E., Morgan, K., Lashley, T., Warner, T. T., Jaunmuktane, Z., Galasko, D., Santana, I., Tienari, P. J., Myllykangas, L., Oinas, M., Cairns, N. J., Morris, J. C., Halliday, G. M., Deerlin, V. M. V., Trojanowski, J. Q., Grassano, M., Calvo, A., Mora, G., Canosa, A., Floris, G., Bohannan, R. C., Brett, F., Gan-Or, Z., Geiger, J. T., Moore, A., May, P., Krüger, R., Goldstein, D. S., Lopez, G., Tayebi, N., Sidransky, E., Sotis, A. R., Sukumar, G., Alba, C., Lott, N., Martinez, E. M., Tuck, M., Singh, J., Bacikova, D., Zhang, X., Hupalo, D. N., Adeleye, A., Wilkerson, M. D., Pollard, H. B., Norcliffe-Kaufmann, L., Palma, J.-A., Kaufmann, H., Shakkottai, V. G., Perkins, M., Newell, K. L., Gasser, T., Schulte, C., Landi, F., Salvi, E., Cusi, D., Masliah, E., Kim, R. C., Caraway, C. A., Monuki, E. S., Brunetti, M., Dawson, T. M., Rosenthal, L. S., Albert, M. S., Pletnikova, O., Troncoso, J. C., Flanagan, M. E., Mao, Q., Bigio, E. H., Rodríguez-Rodríguez, E., Infante, J., Lage, C., González-Aramburu, I., Sanchez-Juan, P., Ghetti, B., Keith, J., Black, S. E., Masellis, M., Rogaeva, E., Duyckaerts, C., Brice, A., Lesage, S., Xiromerisiou, G., Barrett, M. J., Tilley, B. S., Gentleman, S., Logroscino, G., Serrano, G. E., Beach, T. G., McKeith, I. G., Thomas, A. J., Attems, J., Morris, C. M., Palmer, L., Love, S., Troakes, C., Al-Sarraj, S., Hodges, A. K., Aarsland, D., Klein, G., Kaiser, S. M., Woltjer, R., Pastor, P., Bekris, L. M., Leverenz, J. B., Besser, L. M., Kuzma, A., Renton, A. E., Goate, A., Bennett, D. A., Scherzer, C. R., Morris, H. R., Ferrari, R., Albani, D., Pickering-Brown, S., Faber, K., Kukull, W. A., Morenas-Rodriguez, E., Lleó, A., Fortea, J., Alcolea, D., Clarimon, J., Nalls, M. A., Ferrucci, L., Resnick, S. M., Tanaka, T., Foroud, T. M., Graff-Radford, N. R., Wszolek, Z. K., Ferman, T., Boeve, B. F., Hardy, J. A., Topol, E. J., Torkamani, A., Sin-

gleton, A. B., Ryten, M., Dickson, D. W., Chiò, A., Ross, O. A., Gibbs, J. R., Dalgard, C. L., Traynor, B. J., and Scholz, S. W. Genome sequencing analysis identifies new loci associated with lewy body dementia and provides insights into its genetic architecture. *Nature Genetics 53* (3 2021), 294–303. 2.1.2

[53] Christensen, J., Grønborg, T. K., Sørensen, M. J., Schendel, D., Parner, E. T., Pedersen, L. H., and Vestergaard, M. Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA 309* (4 2013), 1696. 4.1.2

[54] Clark, L. N., Kartsaklis, L. A., Gilbert, R. W., Dorado, B., Ross, B. M., Kisselev, S., Verbitsky, M., Mejia-Santana, H., Cote, L. J., Andrews, H., Vonsattel, J.-P., Fahn, S., Mayeux, R., Honig, L. S., and Marder, K. Association of glucocerebrosidase mutations with dementia with lewy bodies. *Archives of neurology 66* (5 2009), 578–83. 2.1.2

[55] Clinton, L. K., Blurton-Jones, M., Myczek, K., Trojanowski, J. Q., and LaFerla, F. M. Synergistic interactions between abeta, tau, and alpha-synuclein: acceleration of neuropathology and cognitive decline. *The Journal of neuroscience : the official journal of the Society for Neuroscience 30* (5 2010), 7281–9. 2.1.1

[56] Cloonan, N., Forrest, A., and Kolle, G. Stem cell transcriptome profiling via massive-scale mrna sequencing. *Nat Methods 5*, 7, 613–619. 2. 1

[57] Compta, Y., Parkkinen, L., Kempster, P., Selikhova, M., Lashley, T., Holton, J. L., Lees, A. J., and Revesz, T. The significance of -synuclein, amyloid- and tau pathologies in parkinson's disease progression and related dementia. *Neurodegenerative Diseases 13* (2014), 154–156. 2.1.1, 1

[58] Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., and Mortazavi, A. A survey of best practices for rna-seq data analysis. *Genome Biology 17* (12 2016), 13. 1.1.2

[59] Consortium, . G. P., et al. An integrated map of genetic variation from 1,092 human genomes. *Nature 491*, 7422 (2012), 56. 2.2.9

[60] Consortium, I. H. ., et al. Integrating common and rare genetic variation in diverse human populations. *Nature 467*, 7311 (2010), 52. 2.2.9

[61] Consortium, T. Creating the gene ontology resource: Design and implementation. *Genome Res 11*, 8, 1425–1433. 43. 1.1.3

[62] Coughlin, D. G., Hurtig, H. I., and Irwin, D. J. Pathological influences on clinical heterogeneity in lewy body diseases. *Movement disorders : official journal of the Movement Disorder Society 35* (1 2020), 5–19. 2.1, 2.1.1

[63] Cui, K., Wang, Y., Zhu, Y., Tao, T., Yin, F., Guo, Y., Liu, H., Li, F., Wang, P., Chen, Y., and Qin, J. Neurodevelopmental impairment induced by prenatal valproic acid exposure shown with the human cortical organoid-on-a-chip model. *Microsystems Nanoengineering 6* (12 2020). 4.1.3, 3, 4.3.7, 4.4

[64] Dang, J., Tiwari, S. K., Agrawal, K., Hui, H., Qin, Y., and Rana, T. M. Glial cell diversity and methamphetamine-induced neuroinflammation in human cerebral organoids. *Molecular Psychiatry 26* (4 2021), 1194–1207. 4.1.2

[65] Daugaard, C. A., Pedersen, L., Sun, Y., Dreier, J. W., and Christensen, J. Association of prenatal exposure to valproate and other antiepileptic drugs with intellectual disability and delayed childhood milestones. *JAMA Network Open 3* (11 2020), e2025570. 4.1.2

[66] Day, J. O., and Mullin, S. The genetics of parkinson's disease and implications for clinical practice. *Genes 12* (6 2021), 1006. 2.1.2

[67] de Leeuw, C. A., Mooij, J. M., Heskes, T., and Posthuma, D. Magma: generalized gene-set analysis of gwas data. *PLoS computational biology 11*, 4 (2015), e1004219. 3.2.11

[68] de Lichtenberg, U., Jensen, L. J., Brunak, S., and Bork, P. Dynamic complex formation during the yeast cell cycle. *science 307*, 5710 (2005), 724–727. 5.1.7

[69] Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Churchhouse, C., Dumont, A., Eriksson, N., Gandal, M., Goldstein, J. I., Grasby, K. L., Grove, J., Gudmundsson, O. O., Hansen, C. S., Hauberg, M. E., Hollegaard, M. V., Howrigan, D. P., Huang, H., Maller, J. B., Martin, A. R., Martin, N. G., Moran, J., Pallesen, J., Palmer, D. S., Pedersen, C. B., Pedersen, M. G., Poterba, T., Poulsen, J. B., Ripke, S., Robinson, E. B., Satterstrom, F. K., Stefansson, H., Stevens, C., Turley, P., Walters, G. B., Won, H., Wright, M. J., Andreassen, O. A., Asherson, P., Burton, C. L., Boomsma, D. I., Cormand, B., Dalsgaard, S., Franke, B., Gelernter, J., Geschwind, D., Hakonarson, H., Haavik, J., Kranzler, H. R., Kuntsi, J., Langley, K., Lesch, K.-P., Middeldorp, C., Reif, A., Rohde, L. A., Roussos, P., Schachar, R., Sklar, P., Sonuga-Barke, E. J. S., Sullivan, P. F., Thapar, A., Tung, J. Y., Waldman, I. D., Medland, S. E., Stefansson, K., Nordentoft, M., Hougaard, D. M., Werge, T., Mors, O., Mortensen, P. B., Daly, M. J., Faraone, S. V., Børglum, A. D., and Neale, B. M. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics 51* (1 2019). 4.2.5, 4.3.5

[70] DeTure, M. A., and Dickson, D. W. The neuropathological diagnosis of alzheimer's disease. *Molecular Neurodegeneration 14* (12 2019), 32. 3.1.1, 3.1.2

[71] DeVos, S. L., Corjuc, B. T., Oakley, D. H., Nobuhara, C. K., Bannon, R. N., Chase, A., Commins, C., Gonzalez, J. A., Dooley, P. M., Frosch, M. P., and Hyman, B. T. Synaptic tau seeding precedes tau pathology in human alzheimer's disease brain. *Frontiers in Neuroscience 12* (4 2018). 3.1.2

[72] Dickson, D. Neuropathological diagnosis of alzheimer's disease: A perspective from longitudinal clinicopathological studies. *Neurobiology of Aging 18* (7 1997), S21–S26. 3.1.2

[73] Divo, M. J., Martinez, C. H., and Mannino, D. M. Ageing and the epidemiology of multimorbidity. *European Respiratory Journal 44* (10 2014), 1055–1068. 5.1.4

[74] Do, J., McKinney, C., Sharma, P., and Sidransky, E. Glucocerebrosidase and its relevance to parkinson disease. *Molecular Neurodegeneration 14* (12 2019), 36. 2.1.2, 2.4.1

[75] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. Star: ultrafast universal rna-seq aligner. *Bioinformatics 29* (1 2013). 1.1.3, 4.2.2

[76] Domínguez-Iturza, N., Lo, A. C., Shah, D., Armendáriz, M., Vannelli, A., Mercaldo, V., Trusel, M., Li, K. W., Gastaldo, D., Santos, A. R., et al. The autism-and schizophrenia-associated protein cyfip1 regulates bilateral brain connectivity and behaviour. *Nature communications 10*, 1 (2019), 3454. 1

[77] Dooley, T., v., C. E., and Reddy, S. Regulation of gene expression in inflammatory bowel disease and correlation with ibd drugs. *Inflamm Bowel Dis 10*, 1, 1–14. 3. 1

[78] Dunckley, T., Beach, T. G., Ramsey, K. E., Grover, A., Mastroeni, D., Walker, D. G., LaFleur, B. J., Coon, K. D., Brown, K. M., Caselli, R., Kukull, W., Higdon, R., McKeel, D., Morris, J. C., Hulette, C., Schmechel, D., Reiman, E. M., Rogers, J., and Stephan, D. A. Gene expression correlates of neurofibrillary tangles in alzheimer's disease. *Neurobiology of Aging 27* (10 2006), 1359–1371. 1, 1

[79] Durinck, S., Spellman, P. T., Birney, E., and Huber, W. Mapping identifiers for the integration of genomic datasets with the r/bioconductor package biomart. *Nature Protocols 4* (8 2009), 1184–1191. 2

[80] Ehlers, K., Elmazar, M. M. A., and Nau, H. Methionine reduces the valproic acid-induced spina bifida rate in mice without altering valproic acid kinetics. *The Journal of Nutrition 126* (1 1996), 67–75. 4.4.5

[81] Emre, M., Aarsland, D., Brown, R., Burn, D. J., Duyckaerts, C., Mizuno, Y., Broe, G. A., Cummings, J., Dickson, D. W., Gauthier, S., Goldman, J., Goetz, C., Korczyn, A., Lees, A., Levy, R., Litvan, I., McKeith, I., Olanow, W., Poewe, W., Quinn, N., Sampaio, C., Tolosa, E., and Dubois, B. Clinical diagnostic criteria for dementia associated with parkinson's disease. *Movement Disorders 22* (9 2007), 1689–1707. 1.2.2, 2.2.1

[82] Engelhardt, E., and da Mota Gomes, M. Lewy and his inclusion bodies: Discovery and rejection. *Dementia Neuropsychologia 11* (6 2017), 198–201. 2.1

[83] Escayg, A., and Goldin, A. Sodium channel scn1a and epilepsy: Mutations and mechanisms. *Epilepsia 51*, 9, 1650–1658. 1.2.1

[84] Extance, A. Alzheimer's failure raises questions about disease-modifying strategies. *Nature Reviews Drug Discovery 9* (10 2010), 749–751. 3.1.5

[85] Fabris, F., de Magalhães, J. P., and Freitas, A. A. A review of supervised machine learning applied to ageing research. *Biogerontology 18* (4 2017), 171–188. 2

[86] Feleke, R., Jazayeri, D., Abouzeid, M., Powell, K. L., Srivastava, P. K., O'Brien, T. J., Jones, N. C., and Johnson, M. R. Integrative genomics reveals pathogenic mediator of valproate-induced neurodevelopmental disability. *Brain 145* (11 2022), 3832–3842. 4.4, 4.5, 4.7

[87] Feleke, R., Reynolds, R. H., Smith, A. M., Tilley, B., Taliun, S. A. G., Hardy, J., Matthews, P. M., Gentleman, S., Owen, D. R., Johnson, M. R., Srivastava, P. K., and Ryten, M. Cross-platform transcriptional profiling identifies common and distinct molecular pathologies in lewy body diseases. *Acta neuropathologica 142* (9 2021), 449–474. 2.3

[88] Finak, G., McDavid, A., and Yajima, M. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna sequencing data. *Genome Biol 16*, 1. 37. 1.1.3

[89] Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S., and Gottardo, R. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell rna sequencing data. *Genome Biology 16* (12 2015), 278. 1.1.3, 2.2.7, 2.4, 3.2.9

[90] Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., Patterson, N., Neale, B. M., and Price, A. L. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics 47* (11 2015), 1228–1235. 2.2.9, 3.2.11

[91] Fleming, S. J., Marioni, J. C., and Babadi, M. Cellbender remove-background: a deep generative model for unsupervised removal of background noise from scrna-seq datasets. *BioRxiv 791699* (2019). 3.2.7

[92] Francés, L., Quintero, J., and Fernández, A. Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the dsm-5: a systematic review in accordance with the prisma criteria. *Child Adolesc Psychiatry Ment Health 16*, 1. 58. 1.2.1, 4.1

[93] Friedman, B., Srinivasan, K., and Ayalon, G. Diverse brain myeloid expression profiles reveal distinct microglial activation states and aspects of alzheimer ' s disease not evident in mouse models resource diverse brain myeloid expression profiles reveal distinct microglial activation states and asp. *CellReports 22*, 3, 832–847. 14. 1.1.1

[94] G, M., K, S., and A, F. Molecular architecture of the developing mouse brain. *Nature 596*, 7870. 283. 4.2.6

[95] Gan, J., Chen, Z., Shi, Z., Li, X., Liu, S., Liu, Y., Zhu, H., Shen, L., Zhang, G., You, Y., Guo, Q., Zhang, N., Lv, Y., Gang, B., Yuan, J., and Ji, Y. Sex differences in clinical cognitive impairment with lewy bodies: a chinese multicenter study. *Biology of Sex Differences 13* (10 2022), 55. 5.1.4

[96] Gandal, M. J., Zhang, P., Hadjimichael, E., Walker, R. L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., Shieh, A. W., Haney, J., Parhami, S., Belmont, J., Kim, M., Losada, P. M., Khan, Z., Mleczko, J., Xia, Y., Dai, R., Wang, D., Yang, Y. T., Xu, M., Fish, K., Hof, P. R., Warrell, J., Fitzgerald, D., White, K., Jaffe, A. E., Peters, M. A., Gerstein, M., Liu, C., Iakoucheva, L. M., Pinto, D., Geschwind, D. H., Ashley-Koch, A. E., Crawford, G. E., Garrett, M. E., Song, L., Safi, A., Johnson, G. D., Wray, G. A., Reddy, T. E., Goes, F. S., Zandi, P., Bryois, J., Jaffe, A. E., Price, A. J., Ivanov, N. A., Collado-Torres, L., Hyde, T. M., Burke, E. E., Kleiman, J. E., Tao, R., Shin, J. H., Akbarian, S., Girdhar, K., Jiang, Y., Kundakovic, M., Brown, L., Kassim, B. S., Park, R. B., Wiseman, J. R., Zharovsky, E., Jacobov, R., Devillers, O., Flatow, E., Hoffman, G. E., Lipska, B. K., Lewis, D. A., Haroutunian, V., Hahn, C.-G., Charney, A. W., Dracheva, S., Kozlenkov, A., Belmont, J., DelValle, D., Francoeur, N., Hadjimichael, E., Pinto, D., van Bakel, H., Roussos, P., Fullard, J. F., Bendl, J., Hauberg, M. E., Mangravite, L. M., Peters, M. A., Chae, Y., Peng, J., Niu, M., Wang, X., Webster, M. J., Beach, T. G., Chen, C., Jiang, Y., Dai, R., Shieh, A. W., Liu, C., Grennan, K. S., Xia, Y., Vadukapuram, R., Wang, Y., Fitzgerald, D., Cheng, L., Brown, M., Brown, M., Brunetti, T., Goodman, T., Alsayed, M., Gandal, M. J., Geschwind, D. H., Won, H., Polioudakis, D., Wamsley, B., Yin, J., Hadzic, T., Ubieta, L. D. L. T., Swarup, V., Sanders, S. J., State, M. W., Werling, D. M., An, J.-Y., Sheppard, B., Willsey, A. J., White, K. P., Ray, M., Giase, G., Kefi, A., Mattei, E., Purcaro, M., Weng, Z., Moore, J., Pratt, H., Huey, J., Borrman, T., Sullivan, P. F., Giusti-Rodriguez, P., Kim, Y., Sullivan, P., Szatkiewicz, J., Rhie,

S. K., Armoskus, C., Camarena, A., Farnham, P. J., Spitsyna, V. N., Witt, H., Schreiner, S., Evgrafov, O. V., Knowles, J. A., Gerstein, M., Liu, S., Wang, D., Navarro, F. C. P., Warrell, J., Clarke, D., Emani, P. S., Gu, M., Shi, X., Xu, M., Yang, Y. T., Kitchen, R. R., Gürsoy, G., Zhang, J., Carlyle, B. C., Nairn, A. C., Li, M., Pochareddy, S., Sestan, N., Skarica, M., Li, Z., Sousa, A. M. M., Santpere, G., Choi, J., Zhu, Y., Gao, T., Miller, D. J., Cherskov, A., Yang, M., Amiri, A., Coppola, G., Mariani, J., Scuderi, S., Szekely, A., Vaccarino, F. M., Wu, F., Weissman, S., Roychowdhury, T., and Abyzov, A. Transcriptome-wide isoform-level dysregulation in asd, schizophrenia, and bipolar disorder. *Science 362* (12 2018). 2

[97] Garcia, A. X., Xu, J., Cheng, F., Ruppin, E., and Schäffer, A. A. Altered gene expression in excitatory neurons is associated with alzheimer's disease and its higher incidence in women. *Alzheimer's Dementia: Translational Research Clinical Interventions 9* (1 2023). 1

[98] Gean, P.-W., Huang, C.-C., Hung, C.-R., and Tsai, J.-J. Valproic acid suppresses the synaptic response mediated by the nmda receptors in rat amygdalar slices. *Brain Research Bulletin 33* (1 1994), 333–336. 4.1.2

[99] Gerson, J., Castillo-Carranza, D. L., Sengupta, U., Bodani, R., Prough, D. S., DeWitt, D. S., Hawkins, B. E., and Kayed, R. Tau oligomers derived from traumatic brain injury cause cognitive impairment and accelerate onset of pathology in htau mice. *Journal of Neurotrauma 33* (11 2016), 2034–2043. 3.1.2

[100] Geut, H., Hepp, D. H., Foncke, E., Berendse, H. W., Rozemuller, J. M., Huitinga, I., and van de Berg, W. D. J. Neuropathological correlates of parkinsonian disorders in a large dutch autopsy series. *Acta Neuropathologica Communications 8* (12 2020), 39. 2.1.1

[101] Goetz, C. The history of parkinson's disease: Early clinical descriptions and neurological therapies. *Cold Spring Harb Perspect Med 1*, 1. 79. 2.1

[102] GOKCAL, E., GUR, V. E., SELVITOP, R., YILDIZ, G. B., AND ASIL, T. Motor and non-motor symptoms in parkinson's disease: Effects on quality of life. *Noro Psikiyatri Arsivi 54* (6 2017), 143–148. 2.1

[103] GOLDMAN, S. M., MAREK, K., OTTMAN, R., MENG, C., COMYNS, K., CHAN, P., MA, J., MARRAS, C., LANGSTON, J. W., ROSS, G. W., AND TANNER, C. M. Concordance for parkinson's disease in twins: A 20-year update. *Annals of Neurology 85* (4 2019), 600–605. 2.1.3

[104] GOMPERTS, S. N. Lewy body dementias: Dementia with lewy bodies and parkinson disease dementia. *Continuum (Minneapolis, Minn.) 22* (4 2016), 435–63. 1

[105] GROVE, J., RIPKE, S., ALS, T. D., MATTHEISEN, M., WALTERS, R. K., WON, H., PALLESEN, J., AGERBO, E., ANDREASSEN, O. A., ANNEY, R., AWASHTI, S., BELLIVEAU, R., BETTELLA, F., BUXBAUM, J. D., BYBJERG-GRAUHOLM, J., BÆKVAD-HANSEN, M., CERRATO, F., CHAMBERT, K., CHRISTENSEN, J. H., CHURCHHOUSE, C., DELLENVALL, K., DEMONTIS, D., RUBEIS, S. D., DEVLIN, B., DJUROVIC, S., DUMONT, A. L., GOLDSTEIN, J. I., HANSEN, C. S., HAUBERG, M. E., HOLLEGAARD, M. V., HOPE, S., HOWRIGAN, D. P., HUANG, H., HULTMAN, C. M., KLEI, L., MALLER, J., MARTIN, J., MARTIN, A. R., MORAN, J. L., NYEGAARD, M., NÆRLAND, T., PALMER, D. S., PALOTIE, A., PEDERSEN, C. B., PEDERSEN, M. G., DPOTERBA, T., POULSEN, J. B., POURCAIN, B. S., QVIST, P., REHNSTRÖM, K., REICHENBERG, A., REICHERT, J., ROBINSON, E. B., ROEDER, K., ROUSSOS, P., SAEMUNDSEN, E., SANDIN, S., SATTERSTROM, F. K., SMITH, G. D., STEFANSSON, H., STEINBERG, S., STEVENS, C. R., SULLIVAN, P. F., TURLEY, P., WALTERS, G. B., XU, X., STEFANSSON, K., GESCHWIND, D. H., NORDENTOFT, M., HOUGAARD, D. M., WERGE, T., MORS, O., MORTENSEN, P. B., NEALE, B. M., DALY, M. J., AND BØRGLUM, A. D. Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics 51* (3 2019). 4.2.5, 4.3.5

[106] GUERREIRO, R., ROSS, O. A., KUN-RODRIGUES, C., HERNANDEZ, D. G., ORME, T., EICHER, J. D., SHEPHERD, C. E., PARKKINEN, L., DARWENT, L., HECKMAN,

M. G., Scholz, S. W., Troncoso, J. C., Pletnikova, O., Ansorge, O., Clarimon, J., Lleo, A., Morenas-Rodriguez, E., Clark, L., Honig, L. S., Marder, K., Lemstra, A., Rogaeva, E., George-Hyslop, P. S., Londos, E., Zetterberg, H., Barber, I., Braae, A., Brown, K., Morgan, K., Troakes, C., Al-Sarraj, S., Lashley, T., Holton, J., Compta, Y., Deerlin, V. V., Serrano, G. E., Beach, T. G., Lesage, S., Galasko, D., Masliah, E., Santana, I., Pastor, P., Diez-Fairen, M., Aguilar, M., Tienari, P. J., Myllykangas, L., Oinas, M., Revesz, T., Lees, A., Boeve, B. F., Petersen, R. C., Ferman, T. J., Escott-Price, V., Graff-Radford, N., Cairns, N. J., Morris, J. C., Pickering-Brown, S., Mann, D., Halliday, G. M., Hardy, J., Trojanowski, J. Q., Dickson, D. W., Singleton, A., Stone, D. J., and Bras, J. Investigating the genetic architecture of dementia with lewy bodies: a two-stage genome-wide association study. *The Lancet. Neurology 17* (1 2018), 64–74. 2.1.3

[107] Guo, P., Gong, W., Li, Y., Liu, L., Yan, R., Wang, Y., Zhang, Y., and Yuan, Z. Pinpointing novel risk loci for lewy body dementia and the shared genetic etiology with alzheimer's disease and parkinson's disease: a large-scale multi-trait association analysis. *BMC Medicine 20* (12 2022), 214. 2.1.2

[108] Gygi, S. P., Rochon, Y., Franza, B. R., and Aebersold, R. Correlation between protein and mrna abundance in yeast. *Molecular and Cellular Biology 19* (3 1999), 1720–1730. 5.1.7

[109] Habela, C. W., Song, H., and li Ming, G. Modeling synaptogenesis in schizophrenia and autism using human ipsc derived neurons. *Molecular and Cellular Neuroscience 73* (6 2016), 52–62. 1

[110] Habib, N., Avraham-Davidi, I., and Basu, A. Massively parallel single-nucleus rna-seq with dronc-seq. *Nat Methods 14*, 10, 955–958. 17. 1.1.1

[111] Habib, N., McCabe, C., Medina, S., Varshavsky, M., Kitsberg, D., Dvir-Szternfeld, R., Green, G., Dionne, D., Nguyen, L., Marshall, J. L., Chen,

F., Zhang, F., Kaplan, T., Regev, A., and Schwartz, M. Disease-associated astrocytes in alzheimer's disease and aging. *Nature Neuroscience 23* (6 2020), 701–706. 5.1.6

[112] Hafemeister, C., and Satija, R. Normalization and variance stabilization of single-cell rna-seq data using regularized negative binomial regression. *Genome Biology 20* (12 2019), 296. 1.1.3, 3.2.8, 1

[113] Haghverdi, L., Lun, A., Morgan, M., and Marioni, J. Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors. *Nat Biotechnol 36*, 5, 421–427. 26. 1.1.3

[114] Haider, S., and Pal, R. Integrated analysis of transcriptomic and proteomic data. *Current Genomics 14* (2 2013), 91–110. 5.1.7

[115] Halbsgut, L. R., Fahim, E., Kapoor, K., Hong, H., and Friedman, L. K. Certain secondary antiepileptic drugs can rescue hippocampal injury following a critical growth period despite poor anticonvulsant activity and cognitive deficits. *Epilepsy Behavior 29* (12 2013), 466–477. 4.1.2

[116] Harvey, P. Domains of cognition and their assessment. *Dialogues Clin Neurosci 21*, 3, 227–237. 54. 1.2

[117] Heather, J., and Chain, B. The sequence of sequencers: The history of sequencing dna. *Genomics 107*, 1, 1–8. 12. 1.1.1

[118] Heilbron, K., Jensen, M. P., Bandres-Ciga, S., Fontanillas, P., Blauwendraat, C., Nalls, M. A., Singleton, A. B., Smith, G. D., Cannon, P., and Noyce, A. J. Unhealthy behaviours and risk of parkinson's disease: A mendelian randomisation study. *Journal of Parkinson's Disease 11* (10 2021), 1981–1993. 2.1.3

[119] Hepp, D. H., Vergoossen, D. L., Huisman, E., Lemstra, A. W., Berendse, H. W., Rozemuller, A. J., Foncke, E. M., and van de Berg, W. D. Distribution and load of amyloid- pathology in parkinson disease and dementia with lewy bodies. *Journal of Neuropathology Experimental Neurology 75* (10 2016), 936–945. 2.1.1

[120] Hoang, T., Wang, J., and Boyd, P. Gene regulatory networks controlling vertebrate retinal regeneration. *Science*, 2020 370 6519. 5. 1

[121] Hollestein, V., Poelmans, G., Forde, N. J., Beckmann, C. F., Ecker, C., Mann, C., Schäfer, T., Moessnang, C., Baumeister, S., Banaschewski, T., Bourgeron, T., Loth, E., Dell'Acqua, F., Murphy, D. G. M., Puts, N. A., Tillmann, J., Charman, T., Jones, E. J. H., Mason, L., Ambrosino, S., Holt, R., Bölte, S., Buitelaar, J. K., and Naaijen, J. Excitatory/inhibitory imbalance in autism: the role of glutamate and gaba gene-sets in symptoms and cortical brain structure. *Translational Psychiatry 13* (1 2023), 18. 2

[122] Hood, L., and Rowen, L. The human genome project: big science transforms biology and medicine. *Genome Med 5*, 9. 10. 1.1

[123] Hook, P., McClymont, S., and Cannon, G. Single-cell rna-seq of mouse dopaminergic neurons informs candidate gene selection for sporadic parkinson disease. *Am J Hum Genet 102*, 3, 427–446. 15. 1.1.1

[124] Hou, Y., Dan, X., Babbar, M., Wei, Y., Hasselbalch, S. G., Croteau, D. L., and Bohr, V. A. Ageing as a risk factor for neurodegenerative disease. *Nature Reviews Neurology 15* (10 2019), 565–581. 1.2.2, 5.1.4

[125] Hsieh, Y.-C., Guo, C., Yalamanchili, H. K., Abreha, M., Al-Ouran, R., Li, Y., Dammer, E. B., Lah, J. J., Levey, A. I., Bennett, D. A., Jager, P. L. D., Seyfried, N. T., Liu, Z., and Shulman, J. M. Tau-mediated disruption of the spliceosome triggers cryptic rna splicing and neurodegeneration in alzheimer's disease. *Cell Reports 29* (10 2019), 301–316.e10. 3, 1

[126] Hsu, Y.-T., Chang, Y.-G., and Chern, Y. Insights into gaba ¡sub¿a¡/sub¿ ergic system alteration in huntington's disease. *Open Biology 8* (12 2018). 2

[127] Hwang, B., Lee, J. H., and Bang, D. Single-cell rna sequencing technologies and bioinformatics pipelines. *Experimental Molecular Medicine 50* (8 2018), 1–14. 1.1.3, 5.1.5

[128] Hyman, B. Tau propagation, different tau phenotypes, and prion-like properties of tau. *Neuron 82* (6 2014), 1189–1190. 3.1.3

[129] Hyman, S. L. Mental retardation, 2007. 1.2.1

[130] Ikeda, M., Kawarai, T., Kawarabayashi, T., Matsubara, E., Murakami, T., Sasaki, A., Tomidokoro, Y., Ikarashi, Y., Kuribara, H., Ishiguro, K., Hasegawa, M., Yen, S.-H., Chishti, M. A., Harigaya, Y., Abe, K., Okamoto, K., George-Hyslop, P. S., Westaway, D., and Shoji†, M. Accumulation of filamentous tau in the cerebral cortex of human tau r406w transgenic mice. *The American Journal of Pathology 166* (2 2005), 521–531. 3.1.2

[131] Irwin, D. J., White, M. T., Toledo, J. B., Xie, S. X., Robinson, J. L., Deerlin, V. V., Lee, V. M.-Y., Leverenz, J. B., Montine, T. J., Duda, J. E., Hurtig, H. I., and Trojanowski, J. Q. Neuropathologic substrates of parkinson disease dementia. *Annals of Neurology 72* (10 2012), 587–598. 2.1.4, 1

[132] Iwaki, H., Blauwendraat, C., Leonard, H. L., Liu, G., Maple-Grødem, J., Corvol, J.-C., Pihlstrøm, L., van Nimwegen, M., Hutten, S. J., Nguyen, K.-D. H., Rick, J., Eberly, S., Faghri, F., Auinger, P., Scott, K. M., Wijeyekoon, R., Deerlin, V. M. V., Hernandez, D. G., Day-Williams, A. G., Brice, A., Alves, G., Noyce, A. J., Tysnes, O.-B., Evans, J. R., Breen, D. P., Estrada, K., Wegel, C. E., Danjou, F., Simon, D. K., Ravina, B., Toft, M., Heutink, P., Bloem, B. R., Weintraub, D., Barker, R. A., Williams-Gray, C. H., van de Warrenburg, B. P., Hilten, J. J. V., Scherzer, C. R., Singleton, A. B., and Nalls, M. A. Genetic risk of parkinson disease and progression:. *Neurology Genetics 5* (8 2019), e348. 2.1.2, 2

[133] Jagadeesh, K. A., Dey, K. K., Montoro, D. T., Mohan, R., Gazal, S., Engreitz, J. M., Xavier, R. J., Price, A. L., and Regev, A. Identifying disease-critical cell types and cellular processes by integrating single-cell rna-sequencing and human genetics. *Nature Genetics 54* (10 2022), 1479–1492. 1

[134] Jansen, I. E., Savage, J. E., Watanabe, K., Bryois, J., Williams, D. M., Steinberg, S., Sealock, J., Karlsson, I. K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I. S., Andersen, F., Bergh, S., Bettella, F., Bjornsson, S., Brækhus, A., Bråthen, G., de Leeuw, C., Desikan, R. S., Djurovic, S., Dumitrescu, L., Fladby, T., Hohman, T. J., Jonsson, P. V., Kiddle, S. J., Rongve, A., Saltvedt, I., Sando, S. B., Selbæk, G., Shoai, M., Skene, N. G., Snaedal, J., Stordal, E., Ulstein, I. D., Wang, Y., White, L. R., Hardy, J., Hjerling-Leffler, J., Sullivan, P. F., van der Flier, W. M., Dobson, R., Davis, L. K., Stefansson, H., Stefansson, K., Pedersen, N. L., Ripke, S., Andreassen, O. A., and Posthuma, D. Genome-wide meta-analysis identifies new loci and functional pathways influencing alzheimer's disease risk. *Nature Genetics 51* (3 2019), 404–413. 2.2.9, 2.3.7

[135] Jazayeri, D., Braine, E., McDonald, S., Dworkin, S., Powell, K. L., Griggs, K., Vajda, F. J. E., O'Brien, T. J., and Jones, N. C. A rat model of valproate teratogenicity from chronic oral treatment during pregnancy. *Epilepsia 61* (6 2020), 1291–1300. 4.1.3, 4.2.1, 4.3.1, 4.3.3, 4.4

[136] Jellinger, K. A. A critical evaluation of current staging of -synuclein pathology in lewy body disorders. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease 1792* (7 2009), 730–740. 2.1.1

[137] Jellinger, K. A., and Korczyn, A. D. Are dementia with lewy bodies and parkinson's disease dementia the same disease? *BMC Medicine 16* (12 2018), 34. 2.1, 2.1.1

[138] Jensen, O. N. Interpreting the protein language using proteomics. *Nature Reviews Molecular Cell Biology 7* (6 2006), 391–403. 5.1.7

[139] Jiang, X., Yu, M., Ou, Y., Cao, Y., Yao, Y., Cai, P., and Zhang, F. Downregulation of usp4 promotes activation of microglia and subsequent neuronal inflammation in rat spinal cord after injury. *Neurochemical Research 42* (11 2017), 3245–3253. 2

[140] Jicha, G. A., Berenfeld, B., and Davies, P. Sequence requirements for formation of conformational variants of tau similar to those found in alzheimer's disease. *Journal of Neuroscience Research 55* (3 1999), 713–723. 3.1.4

[141] Jiwaji, Z., Tiwari, S. S., Avilés-Reyes, R. X., Hooley, M., Hampton, D., Torvell, M., Johnson, D. A., McQueen, J., Baxter, P., Sabari-Sankar, K., Qiu, J., He, X., Fowler, J., Febery, J., Gregory, J., Rose, J., Tulloch, J., Loan, J., Story, D., McDade, K., Smith, A. M., Greer, P., Ball, M., Kind, P. C., Matthews, P. M., Smith, C., Dando, O., Spires-Jones, T. L., Johnson, J. A., Chandran, S., and Hardingham, G. E. Reactive astrocytes acquire neuroprotective as well as deleterious signatures in response to tau and aß pathology. *Nature Communications 13* (1 2022), 135. 5.1.4

[142] Johnson, M. R., Shkura, K., Langley, S. R., Delahaye-Duriez, A., Srivastava, P., Hill, W. D., Rackham, O. J. L., Davies, G., Harris, S. E., Moreno-Moral, A., Rotival, M., Speed, D., Petrovski, S., Katz, A., Hayward, C., Porteous, D. J., Smith, B. H., Padmanabhan, S., Hocking, L. J., Starr, J. M., Liewald, D. C., Visconti, A., Falchi, M., Bottolo, L., Rossetti, T., Danis, B., Mazzuferi, M., Foerch, P., Grote, A., Helmstaedter, C., Becker, A. J., Kaminski, R. M., Deary, I. J., and Petretto, E. Systems genetics identifies a convergent gene network for cognition and neurodevelopmental disease. *Nature Neuroscience 19* (2 2016), 223–232. 1, 4.3.5, 4.3.5, 4.4.3

[143] Jolliffe, I. T., and Cadima, J. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374* (4 2016), 20150202. 1.1.3

[144] K, B., Bézieux H, R., and K, S. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun 11*, 1. 50. 1.1.3

[145] Kalaitzakis, M. E., Gentleman, S. M., and Pearce, R. K. B. Disturbed sleep in parkinson's disease: anatomical and pathological correlates. *Neuropathology and Applied Neurobiology 39* (10 2013), 644–653. 2.1.1

[146] Kalra, S., Bergeron, C., and Lang, A. E. Lewy body disease and dementia. a review. *Archives of internal medicine 156* (3 1996), 487–93. 2.1

[147] Kanaan, N., Morfini, G., and LaPointe, N. Pathogenic forms of tau inhibit kinesin-dependent axonal transport through a mechanism involving activation of axonal phosphotransferases. *Journal of Neuroscience 31*, 27, 9858–9868. 195. 3.1.2

[148] Karmiloff-Smith, A., Casey, B., Massand, E., Tomalski, P., and Thomas, M. Environmental and genetic influences on neurocognitive development. *Clinical Psychological Science 2*, 5, 628–637. 1.2

[149] Kaufman, S. K., Sanders, D. W., Thomas, T. L., Ruchinskas, A. J., Vaquer-Alicea, J., Sharma, A. M., Miller, T. M., and Diamond, M. I. Tau prion strains dictate patterns of cell pathology, progression rate, and regional vulnerability in vivo. *Neuron 92* (11 2016), 796–812. 3.1.3

[150] Kaufman, S. K., Tredici, K. D., Thomas, T. L., Braak, H., and Diamond, M. I. Tau seeding activity begins in the transentorhinal/entorhinal regions and anticipates phospho-tau pathology in alzheimer's disease and part. *Acta Neuropathologica 136* (7 2018), 57–67. 3.1.5

[151] Keller, M. F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Simon-Sanchez, J., Mittag, F., Buchel, F., Sharma, M., Gibbs, J. R., Schulte, C., Moskvina, V., Durr, A., Holmans, P., Kilarski, L. L., Guerreiro, R., Hernandez, D. G., Brice, A., Ylikotila, P., Stefansson, H., Majamaa, K., Morris, H. R., Williams, N., Gasser, T., Heutink, P., Wood, N. W., Hardy, J., Martinez, M., Singleton, A. B., and Nalls, M. A. Using genome-wide complex trait analysis to quantify 'missing heritability' in parkinson's disease. *Human Molecular Genetics 21* (11 2012), 4996–5009. 2.1.3

[152] Kellogg, M., and Meador, K. J. Neurodevelopmental effects of antiepileptic drugs. *Neurochemical Research 42* (7 2017), 2065–2070. 4.4.5

[153] Kharchenko, P. V., Silberstein, L., and Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nature methods 11*, 7 (2014), 740–742. 1.1.3

[154] Kim, H., Kim, Y., Lee, C.-Y., Kim, D.-G., and Cheon, M. Investigation of early molecular alterations in tauopathy with generative adversarial networks. *Scientific Reports 13* (1 2023), 732. 3.1.5

[155] Kim, J.-H., Shinde, D., Reijnders, M., Hauser, N., Belmonte, R., Wilson, G., Bosch, D., Bubulya, P., Shashi, V., Petrovski, S., Stone, J., Park, E., Veltman, J., Sinnema, M., Stumpel, C., Draaisma, J., Nicolai, J., Yntema, H., Lindstrom, K., de Vries, B., Jewett, T., Santoro, S., Vogt, J., Bachman, K., Seeley, A., Krokosky, A., Turner, C., Rohena, L., Hempel, M., Kortüm, F., Lessel, D., Neu, A., Strom, T., Wieczorek, D., Bramswig, N., Laccone, F., Behunova, J., Rehder, H., Gordon, C., Rio, M., Romana, S., Tang, S., El-Khechen, D., Cho, M., McWalter, K., Douglas, G., Baskin, B., Begtrup, A., Funari, T., Schoch, K., Stegmann, A., Stevens, S., Zhang, D.-E., Traver, D., Yao, X., MacArthur, D., Brunner, H., Mancini, G., Myers, R., Owen, L., Lim, S.-T., Stachura, D., Vissers, L., and Ahn, E.-Y. De novo mutations in son disrupt rna splicing of genes essential for brain development and metabolism, causing an intellectual-disability syndrome. *The American Journal of Human Genetics 99* (9 2016), 711–719. 4.1.1

[156] Kim, Y. S., and Yoon, B.-E. Altered gabaergic signaling in brain disease at various stages of life. *Experimental neurobiology 26* (6 2017), 122–131. 4.1.2, 2

[157] Klein, C., and Westenberger, A. Genetics of parkinson's disease. *Cold Spring Harbor perspectives in medicine 2* (1 2012), a008888. 2.1.2

[158] Kong, S. M., Chan, B. K., Park, J.-S., Hill, K. J., Aitken, J. B., Cottle, L., Farghaian, H., Cole, A. R., Lay, P. A., Sue, C. M., and Cooper, A. A. Parkinson's disease-linked human park9/atp13a2 maintains zinc homeostasis and promotes -synuclein externalization via exosomes. *Human Molecular Genetics 23* (6 2014), 2816–2833. 2.4.3

[159] Kouli, A., Torsney, K., and Kuan, W. Parkinson's disease: Etiology, neuropathology, and pathogenesis. *Parkinson's Disease: Pathogenesis and Clinical Aspects*. 2.1.1

[160] Krishnaswami, S. R., Grindberg, R. V., Novotny, M., Venepally, P., Lacar, B., Bhutani, K., Linker, S. B., Pham, S., Erwin, J. A., Miller, J. A., Hodge, R., McCarthy, J. K., Kelder, M., McCorrison, J., Aevermann, B. D., Fuertes, F. D., Scheuermann, R. H., Lee, J., Lein, E. S., Schork, N., McConnell, M. J., Gage, F. H., and Lasken, R. S. Using single nuclei for rna-seq to capture the transcriptome of postmortem neurons. *Nature Protocols 11* (3 2016), 499–524. 2.2.1, 5.1.5

[161] Kritsilis, M., Rizou, S. V., Koutsoudaki, P., Evangelou, K., Gorgoulis, V., and Papadopoulos, D. Ageing, cellular senescence and neurodegenerative disease. *International Journal of Molecular Sciences 19* (9 2018), 2937. 5.1.4

[162] Kukurba, K. R., and Montgomery, S. B. Rna sequencing and analysis. *Cold Spring Harbor Protocols 2015* (11 2015), pdb.top084970. 1.1.2

[163] Kulaga, A. Y., Ursu, E., Toren, D., Tyshchenko, V., Guinea, R., Pushkova, M., Fraifeld, V. E., and Tacutu, R. Machine learning analysis of longevity-associated gene expression landscapes in mammals. *International Journal of Molecular Sciences 22* (1 2021), 1073. 2

[164] La Manno, G., Siletti, K., Furlan, A., Gyllborg, D., Vinsland, E., Mossi Albiach, A., Mattsson Langseth, C., Khven, I., Lederer, A. R., Dratva, L. M., et al. Molecular architecture of the developing mouse brain. *Nature 596*, 7870 (2021), 92–96. 4.2.6

[165] Lacar, B., Linker, S., and Jaeger, B. Nuclear rna-seq of single neurons reveals molecular signatures of activation. *Nat Commun 7*, 11022. 18. 1.1.1

[166] Lake, B. B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., Wildberg, A., Gao, D., Fung, H.-L., Chen, S., Vijayaraghavan, R., Wong, J., Chen, A., Sheng, X., Kaper, F., Shen, R., Ronaghi, M., Fan, J.-B., Wang,

W., Chun, J., and Zhang, K. Neuronal subtypes and diversity revealed by single-nucleus rna sequencing of the human brain. *Science 352* (6 2016), 1586–1590. 1.1.1, 5.1.6

[167] Lake, B. B., Chen, S., Hoshi, M., Plongthongkum, N., Salamon, D., Knoten, A., Vijayan, A., Venkatesh, R., Kim, E. H., Gao, D., Gaut, J., Zhang, K., and Jain, S. A single-nucleus rna-sequencing pipeline to decipher the molecular anatomy and pathophysiology of human kidneys. *Nature Communications 10* (6 2019), 2832. 5.1.5

[168] Landrigan, P., Sonawane, B., Butler, R., Trasande, L., Callan, R., and Droller, D. Early environmental origins of neurodegenerative disease in later life. *Environ Health Perspect 113*, 9, 1230–1233. 69. 1.2.2

[169] Langston, J. W., Ballard, P., Tetrud, J. W., and Irwin, I. Chronic parkinsonism in humans due to a product of meperidine-analog synthesis. *Science 219* (2 1983), 979–980. 2.1.3

[170] Larkby, C., and Day, N. The effects of prenatal alcohol exposure. *Alcohol health and research world 21* (1997), 192–8. 4.1.2

[171] Ledergerber, C., and Dessimoz, C. Base-calling for next-generation sequencing platforms. *Briefings in bioinformatics 12*, 5 (2011), 489–497. 1.1.1

[172] Lee, B.-H., Lee, M. J., Park, S., Oh, D.-C., Elsasser, S., Chen, P.-C., Gartner, C., Dimova, N., Hanna, J., Gygi, S. P., Wilson, S. M., King, R. W., and Finley, D. Enhancement of proteasome activity by a small-molecule inhibitor of usp14. *Nature 467* (9 2010), 179–184. 2

[173] Lee, K., and Lim, C.-Y. Mendelian randomization analysis in observational epidemiology. *Journal of Lipid and Atherosclerosis 8* (2019), 67. 2.1.3, 3

[174] Lee, P. H., Anttila, V., Won, H., Feng, Y.-C. A., Rosenthal, J., Zhu, Z., Tucker-Drob, E. M., Nivard, M. G., Grotzinger, A. D., Posthuma, D., et al. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell 179*, 7 (2019), 1469–1482. 4.2.5, 4.3.5

[175] Lee, R. G., Sedghi, M., Salari, M., Shearwood, A.-M. J., Stentenbach, M., Kariminejad, A., Goullee, H., Rackham, O., Laing, N. G., Tajsharghi, H., and Filipovska, A. Early-onset parkinson disease caused by a mutation in chchd2 and mitochondrial dysfunction. *Neurology Genetics 4* (10 2018), e276. 2.1.2, 4

[176] Lee, S., Abecasis, G. R., Boehnke, M., and Lin, X. Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics 95*, 1 (2014), 5–23. 1.1.3

[177] Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T.-M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., Dee, N. R., Desaki, A. L., Desta, T., Diep, E., Dolbeare, T. A., Donelan, M. J., Dong, H.-W., Dougherty, J. G., Duncan, B. J., Ebbert, A. J., Eichele, G., Estin, L. K., Faber, C., Facer, B. A., Fields, R., Fischer, S. R., Fliss, T. P., Frensley, C., Gates, S. N., Glattfelder, K. J., Halverson, K. R., Hart, M. R., Hohmann, J. G., Howell, M. P., Jeung, D. P., Johnson, R. A., Karr, P. T., Kawal, R., Kidney, J. M., Knapik, R. H., Kuan, C. L., Lake, J. H., Laramee, A. R., Larsen, K. D., Lau, C., Lemon, T. A., Liang, A. J., Liu, Y., Luong, L. T., Michaels, J., Morgan, J. J., Morgan, R. J., Mortrud, M. T., Mosqueda, N. F., Ng, L. L., Ng, R., Orta, G. J., Overly, C. C., Pak, T. H., Parry, S. E., Pathak, S. D., Pearson, O. C., Puchalski, R. B., Riley, Z. L., Rockett, H. R., Rowland, S. A., Royall, J. J., Ruiz, M. J., Sarno, N. R., Schaffnit, K., Shapovalova, N. V., Sivisay, T., Slaughterbeck, C. R., Smith, S. C., Smith, K. A., Smith, B. I., Sodt, A. J., Stewart, N. N., Stumpf, K.-R., Sunkin, S. M., Sutram, M., Tam, A., Teemer, C. D., Thaller, C., Thompson, C. L., Varnam, L. R., Visel, A., Whitlock, R. M., Wohnoutka, P. E., Wolkey, C. K., Wong, V. Y., Wood, M., Yaylaoglu, M. B., Young, R. C., Youngstrom, B. L., Yuan, X. F., Zhang, B., Zwingman, T. A., and Jones, A. R. Genome-

wide atlas of gene expression in the adult mouse brain. *Nature 445* (1 2007), 168–176. 1.1

[178] Leverenz, J. B., Hamilton, R., Tsuang, D. W., Schantz, A., Vavrek, D., Larson, E. B., Kukull, W. A., Lopez, O., Galasko, D., Masliah, E., Kaye, J., Woltjer, R., Clark, C., Trojanowski, J. Q., and Montine, T. J. Research article: Empiric refinement of the pathologic assessment of lewy-related pathology in the dementia patient. *Brain Pathology 18* (1 2008), 220–224. 2.1.1

[179] Li, B., and Leal, S. M. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics 83*, 3 (2008), 311–321. 1.1.3

[180] Li, K., Li, J., Zheng, J., and Qin, S. Reactive astrocytes in neurodegenerative diseases. *Aging and disease 10* (2019), 664. 3.3.6, 3.4.3, 5.1.4

[181] Li, X., and Wang, C.-Y. From bulk, single-cell to spatial rna sequencing. *International Journal of Oral Science 13* (12 2021), 36. 1.1.1, 5.1.5

[182] Liao, Y., Wang, J., Jaehnig, E. J., Shi, Z., and Zhang, B. Webgestalt 2019: gene set analysis toolkit with revamped uis and apis. *Nucleic Acids Research 47* (7 2019), W199–W205. 1.1.3, 2.2.8, 3.2.10, 3.3.4, 4.2.4

[183] Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S., and Kluger, Y. Fast interpolation-based t-sne for improved visualization of single-cell rna-seq data. *Nature Methods 16* (3 2019), 243–245. 1.1.3

[184] Liu, K., Zhao, J., Chen, C., Xu, J., Bell, R. L., Hall, F. S., Koob, G. F., Volkow, N. D., Qing, H., and Lin, Z. Epistatic evidence for gender-dependant slow neurotransmission signalling in substance use disorders: Ppp1r12b versus ppp1r1b. *EBioMedicine 61* (11 2020), 103066. 2.3.7

[185] Liu, X., and Moussa, C. Regulatory role of ubiquitin specific protease-13 (usp13) in misfolded protein clearance in neurodegenerative diseases. *Neuroscience 460* (4 2021), 161–166. 2

[186] Liu, X., Ye, M., and Ma, L. The emerging role of autophagy and mitophagy in tauopathies: From pathogenesis to translational implications in alzheimer's disease. *Frontiers in Aging Neuroscience 14* (10 2022). 1

[187] Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., Brayne, C., Burns, A., Cohen-Mansfield, J., Cooper, C., Costafreda, S. G., Dias, A., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Ogunniyi, A., Orgeta, V., Ritchie, K., Rockwood, K., Sampson, E. L., Samus, Q., Schneider, L. S., Selbæk, G., Teri, L., and Mukadam, N. Dementia prevention, intervention, and care: 2020 report of the lancet commission. *The Lancet 396* (8 2020), 413–446. 2.1.3

[188] Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods 15* (12 2018), 1053–1058. 2

[189] Louros, S. R., and Osterweil, E. K. Perturbed proteostasis in autism spectrum disorders. *Journal of Neurochemistry 139* (12 2016), 1081–1092. 1

[190] Luecken, M., and Theis, F. Current best practices in single-cell rna-seq analysis: a tutorial. *Mol Syst Biol 15*, 6. 25. 1.1.3

[191] Lun, A., Riesenfeld, S., Andrews, T., Dao, T., Gomes, T., and Marioni, J. Emptydrops: distinguishing cells from empty droplets in droplet-based single-cell rna sequencing data. *Genome Biol 20*, 1. 144. 1.1.3

[192] Lun, A. T., McCarthy, D. J., and Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell rna-seq data with bioconductor. *F1000Research 5* (10 2016), 2122. 1.1.3

[193] Lun, A. T. L., Riesenfeld, S., Andrews, T., Dao, T. P., Gomes, T., and Marioni, J. C. Emptydrops: distinguishing cells from empty droplets in droplet-based single-cell rna sequencing data. *Genome Biology 20* (12 2019), 63. 2.2.3, 6

[194] Luo, X.-G., Ding, J.-Q., and Chen, S.-D. Microglia in the aging brain: relevance to neurodegeneration. *Molecular Neurodegeneration 5* (12 2010), 12. 5.1.4

[195] Ma, H., Qiu, R., Zhang, W., Chen, X., Zhang, L., and Wang, M. Association of ppp1r1b polymorphisms with working memory in healthy han chinese adults. *Frontiers in Neuroscience 16* (11 2022). 2.3.7

[196] Macosko, E., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A., Kamitaki, N., Martersteck, E., Trombetta, J., Weitz, D., Sanes, J., Shalek, A., Regev, A., and McCarroll, S. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell 161* (5 2015), 1202–1214. 1.1.3

[197] Maia, N., Nabais Sá, M., Melo-Pires, M., Brouwer, A., and Jorge, P. Intellectual disability genomics: current state, pitfalls and future challenges. *BMC Genomics 22*, 1. 1.2.1

[198] Marescaux, C., Vergnes, M., and Depaulis, A. Genetic absence epilepsy in rats from strasbourg — a review. In *Generalized Non-Convulsive Epilepsy: Focus on GABA-B Receptors*. Springer Vienna. 289. 4.3.1

[199] Mariani, J., Coppola, G., Zhang, P., Abyzov, A., Provini, L., Tomasini, L., Amenduni, M., Szekely, A., Palejev, D., Wilson, M., Gerstein, M., Grigorenko, E. L., Chawarska, K., Pelphrey, K. A., Howe, J. R., and Vaccarino, F. M. Foxg1-dependent dysregulation of gaba/glutamate neuron differentiation in autism spectrum disorders. *Cell 162* (7 2015), 375–390. 2

[200] Marini, C., Porro, A., Rastetter, A., Dalle, C., Rivolta, I., Bauer, D., Oegema, R., Nava, C., Parrini, E., Mei, D., Mercer, C., Dhamija, R., Chambers, C., Coubes, C., Thévenon, J., Kuentz, P., Julia, S., Pasquier, L., Dubourg, C., Carré, W., Rosati, A., Melani, F., Pisano, T., Giardino, M., Innes, A. M., Alembik, Y., Scheidecker, S., Santos, M., Figueiroa, S., Garrido, C., Fusco, C., Frattini, D., Spagnoli, C., Binda, A., Granata, T.,

Ragona, F., Freri, E., Franceschetti, S., Canafoglia, L., Castellotti, B., Gellera, C., Milanesi, R., Mancardi, M. M., Clark, D. R., Kok, F., Helbig, K. L., Ichikawa, S., Sadler, L., Neupauerová, J., Laššuthova, P., Štěrbová, K., Laridon, A., Brilstra, E., Koeleman, B., Lemke, J. R., Zara, F., Striano, P., Soblet, J., Smits, G., Deconinck, N., Barbuti, A., DiFrancesco, D., LeGuern, E., Guerrini, R., Santoro, B., Hamacher, K., Thiel, G., Moroni, A., DiFrancesco, J. C., and Depienne, C. ¡i¿hcn1¡/i¿ mutation spectrum: from neonatal epileptic encephalopathy to benign generalized epilepsy and beyond. *Brain 141* (11 2018), 3160–3178. 4.4.3

[201] Mark, R. J., Ashford, J. W., Goodman, Y., and Mattson, M. P. Anticonvulsants attenuate amyloid -peptide neurotoxicity, ca2+ deregulation, and cytoskeletal pathology. *Neurobiology of Aging 16* (3 1995), 187–198. 4.1.2

[202] Marson, A., Burnside, G., Appleton, R., Smith, D., Leach, J. P., Sills, G., Tudur-Smith, C., Plumpton, C., Hughes, D. A., Williamson, P., Baker, G. A., Balabanova, S., Taylor, C., Brown, R., Hindley, D., Howell, S., Maguire, M., Mohanraj, R., Smith, P. E., Lanyon, K., Manford, M., Chitre, M., Parker, A., Swiderska, N., Appleton, R., Pauling, J., Hughes, A., Gupta, R., Hanif, S., Awadh, M., Ragunathan, S., Cable, N., Cooper, P., Hindley, D., Rakshi, K., Molloy, S., Reuber, M., Ayonrinde, K., Wilson, M., Saladi, S., Gibb, J., Funston, L.-A., Cassidy, D., Boyd, J., Ratnayaka, M., Faza, H., Sadler, M., Al-Moasseb, H., Galtrey, C., Wren, D., Olabi, A., Fuller, G., Khan, M., Kallappa, C., Chinthapalli, R., Aji, B., Davies, R., Foster, K., Hitiris, N., Maguire, M., Hussain, N., Dowson, S., Ellison, J., Sharrack, B., Gandhi, V., Powell, R., Tittensor, P., Summers, B., Shashikiran, S., Dison, P. J., Samarasekera, S., McCorry, D., White, K., Nithi, K., Richardson, M., Brown, R., Page, R., Deekollu, D., Slaght, S., Warriner, S., Ahmed, M., Chaudhuri, A., Chow, G., Artal, J., Kucinskiene, D., Sreenivasa, H., , S. V., Zipitis, C. S., McLean, B., Lal, V., Gregoriou, A., Maddison, P., Pickersgill, T., Anderson, J., Lawthom, C.,

Howell, S., Whitlingum, G., Rakowicz, W., Kinton, L., McLellan, A., Zuberi, S., Kelso, A., Hughes, I., Martland, J., Emsley, H., de Goede, C., Singh, R., Moor, C.-C., Aram, J., Mohanraj, R., Sakthivel, K., Nelapatla, S., Rittey, C., Pinto, A., Leach, J. P., Cock, H., Richardson, A., Houston, E., Cooper, C., Lawson, G., Massarano, A., Burness, C., Marson, A., Smith, D., Wieshmann, U., Dey, I., Sivakumar, P., Yeung, L.-K., Smith, P., Bentur, H., Heafield, T., Mathew, A., Smith, D., and Jauhari, P. The sanad ii study of the effectiveness and cost-effectiveness of valproate versus levetiracetam for newly diagnosed generalised and unclassifiable epilepsy: an open-label, non-inferiority, multicentre, phase 4, randomised controlled trial. *The Lancet 397* (4 2021), 1375–1386. 4.1.2

[203] Marson, A. G., Al-Kharusi, A. M., Alwaidh, M., Appleton, R., Baker, G. A., Chadwick, D. W., Cramp, C., Cockerell, O. C., Cooper, P. N., Doughty, J., Eaton, B., Gamble, C., Goulding, P. J., Howell, S. J., Hughes, A., Jackson, M., Jacoby, A., Kellett, M., Lawson, G. R., Leach, J. P., Nicolaides, P., Roberts, R., Shackley, P., Shen, J., Smith, D. F., Smith, P. E., Smith, C. T., Vanoli, A., and Williamson, P. R. The sanad study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. *The Lancet 369* (3 2007), 1016–1026. 4.1.2, 4.4

[204] Mathys, H., Davila-Velderrain, J., and Peng, Z. Single-cell transcriptomic analysis of alzheimer's disease. *Nature 570*, 7761, 332–337. 1

[205] Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D. A., Kellis, M., and Tsai, L.-H. Single-cell transcriptomic analysis of alzheimer's disease. *Nature 570* (6 2019), 332–337. 1

[206] Mazzulli, J., Xu, Y.-H., Sun, Y., Knight, A., McLean, P., Caldwell, G., Sidransky, E., Grabowski, G., and Krainc, D. Gaucher disease glucocerebrosidase

and -synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell 146* (7 2011), 37–52. 2.4.1

[207] McDermott, S., Wu, J., Cai, B., Lawson, A., and Marjorie Aelion, C. Probability of intellectual disability is associated with soil concentrations of arsenic and lead. *Chemosphere 84*, 1, 31–38. 65. 1.2.1

[208] McDonald, W. Overview of neurocognitive disorders. *Focus (Madison 15*, 1, 4–12. 1.2

[209] McGinnis, C. S., Murrow, L. M., and Gartner, Z. J. Doubletfinder: Doublet detection in single-cell rna sequencing data using artificial nearest neighbors. *Cell Systems 8* (4 2019), 329–337.e4. 1.1.3, 2.2.4, 3.2.7

[210] McKeith, I., Boeve, B., and Dickson, D. Diagnosis and management of dementia with lewy bodies. *Neurology 89*, 1, 88–100. et al. 1.2.2

[211] McKeith, I. G., Boeve, B. F., Dickson, D. W., Halliday, G., Taylor, J.-P., Weintraub, D., Aarsland, D., Galvin, J., Attems, J., Ballard, C. G., Bayston, A., Beach, T. G., Blanc, F., Bohnen, N., Bonanni, L., Bras, J., Brundin, P., Burn, D., Chen-Plotkin, A., Duda, J. E., El-Agnaf, O., Feldman, H., Ferman, T. J., ffytche, D., Fujishiro, H., Galasko, D., Goldman, J. G., Gomperts, S. N., Graff-Radford, N. R., Honig, L. S., Iranzo, A., Kantarci, K., Kaufer, D., Kukull, W., Lee, V. M., Leverenz, J. B., Lewis, S., Lippa, C., Lunde, A., Masellis, M., Masliah, E., McLean, P., Mollenhauer, B., Montine, T. J., Moreno, E., Mori, E., Murray, M., O'Brien, J. T., Orimo, S., Postuma, R. B., Ramaswamy, S., Ross, O. A., Salmon, D. P., Singleton, A., Taylor, A., Thomas, A., Tiraboschi, P., Toledo, J. B., Trojanowski, J. Q., Tsuang, D., Walker, Z., Yamada, M., and Kosaka, K. Diagnosis and management of dementia with lewy bodies. *Neurology 89* (7 2017), 88–100. 2.1.1

[212] Meador, K. J., Baker, G. A., Browning, N., Cohen, M. J., Bromley, R. L., Clayton-Smith, J., Kalayjian, L. A., Kanner, A., Liporace, J. D., Pennell,

P. B., Privitera, M., and Loring, D. W. Fetal antiepileptic drug exposure and cognitive outcomes at age 6 years (nead study): a prospective observational study. *The Lancet Neurology 12* (3 2013), 244–252. 4.1.2

[213] Meeus, B., Verstraeten, A., Crosiers, D., Engelborghs, S., den Broeck, M. V., Mattheijssens, M., Peeters, K., Corsmit, E., Elinck, E., Pickut, B., Vandenberghe, R., Cras, P., Deyn, P. P. D., Broeckhoven, C. V., and Theuns, J. Dlb and pdd: a role for mutations in dementia and parkinson disease genes? *Neurobiology of Aging 33* (3 2012), 629.e5–629.e18. 2.1.2

[214] Meisler, M. H., Hill, S. F., and Yu, W. Sodium channelopathies in neurodevelopmental disorders. *Nature Reviews Neuroscience 22* (3 2021), 152–166. 4.4.3

[215] Meldrum, B. S., and Rogawski, M. A. Molecular targets for antiepileptic drug development. *Neurotherapeutics 4* (1 2007), 18–61. 4.1.2, 1, 4

[216] Meng, Q., Zhang, W., Wang, X., Jiao, C., Xu, S., Liu, C., Tang, B., and Chen, C. Human forebrain organoids reveal connections between valproic acid exposure and autism risk. *Translational Psychiatry 12* (3 2022), 130. 4.1.2

[217] Meredith, G. E., and Rademacher, D. J. Mptp mouse models of parkinson's disease: an update. *Journal of Parkinson's disease 1* (2011), 19–33. 2.1.3

[218] Mhyre, T., Boyd, J., Hamill, R., and Maguire-Zeiss, K. Parkinson's disease. *In 2012*, 389–455. 1.2.2

[219] Monfrini, E., Spagnolo, F., Canesi, M., Seresini, A., Rini, A., Passarella, B., Percetti, M., Seia, M., Goldwurm, S., Cereda, V., Comi, G. P., Pezzoli, G., and Fonzo, A. D. Vps13c-associated parkinson's disease: Two novel cases and review of the literature. *Parkinsonism Related Disorders 94* (1 2022), 37–39. 2.1.2

[220] Murphy, A. E., and Skene, N. G. A balanced measure shows superior performance of pseudobulk methods in single-cell rna-sequencing analysis. *Nature Communications 13* (12 2022), 7851. 5.1.8

[221] Murphy, K. E., Cottle, L., Gysbers, A. M., Cooper, A. A., and Halliday, G. M. Atp13a2 (park9) protein levels are reduced in brain tissue of cases with lewy bodies. *Acta Neuropathologica Communications 1* (12 2013), 11. 2.4.3

[222] Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D. A., Noyce, A. J., Xue, A., Bras, J., Young, E., von Coelln, R., Simón-Sánchez, J., Schulte, C., Sharma, M., Krohn, L., Pihlstrøm, L., Siitonen, A., Iwaki, H., Leonard, H., Faghri, F., Gibbs, J. R., Hernandez, D. G., Scholz, S. W., Botia, J. A., Martinez, M., Corvol, J.-C., Lesage, S., Jankovic, J., Shulman, L. M., Sutherland, M., Tienari, P., Majamaa, K., Toft, M., Andreassen, O. A., Bangale, T., Brice, A., Yang, J., Gan-Or, Z., Gasser, T., Heutink, P., Shulman, J. M., Wood, N. W., Hinds, D. A., Hardy, J. A., Morris, H. R., Gratten, J., Visscher, P. M., Graham, R. R., Singleton, A. B., Adarmes-Gómez, A. D., Aguilar, M., Aitkulova, A., Akhmetzhanov, V., Alcalay, R. N., Alvarez, I., Alvarez, V., Bandres-Ciga, S., Barrero, F. J., Yarza, J. A. B., Bernal-Bernal, I., Billingsley, K., Blauwendraat, C., Blazquez, M., Bonilla-Toribio, M., Botía, J. A., Boungiorno, M. T., Bras, J., Brice, A., Brockmann, K., Bubb, V., Buiza-Rueda, D., Cámara, A., Carrillo, F., Carrión-Claro, M., Cerdan, D., Chelban, V., Clarimón, J., Clarke, C., Compta, Y., Cookson, M. R., Corvol, J.-C., Craig, D. W., Danjou, F., Diez-Fairen, M., Dols-Icardo, O., Duarte, J., Duran, R., Escamilla-Sevilla, F., Escott-Price, V., Ezquerra, M., Faghri, F., Feliz, C., Fernández, M., Fernández-Santiago, R., Finkbeiner, S., Foltynie, T., Gan-Or, Z., Garcia, C., García-Ruiz, P., Gasser, T., Gibbs, J. R., Heredia, M. J. G., Gómez-Garre, P., González, M. M., Gonzalez-Aramburu, I., Guelfi, S., Guerreiro, R., Hardy, J., Hassin-Baer, S., Hernandez, D. G., Heutink, P., Hoenicka, J., Holmans, P., Houlden, H., Infante, J., Iwaki, H., Jesús, S., Jimenez-Escrig, A., Kaishybayeva, G., Kaiyrzhanov, R., Karimova, A., Kia, D. A., Kinghorn, K. J., Koks, S., Krohn, L., Kulisevsky, J., Labrador-Espinosa, M. A., Leonard,

H. L., Lesage, S., Lewis, P., Lopez-Sendon, J. L., Lovering, R., Lubbe, S., Lungu, C., Macias, D., Majamaa, K., Manzoni, C., Marín, J., Marinus, J., Marti, M. J., Martinez, M., Torres, I. M., Martínez-Castrillo, J. C., Mata, M., Mencacci, N. E., del Barrio, C. M., Middlehurst, B., Mínguez, A., Mir, P., Mok, K. Y., Morris, H. R., Muñoz, E., Nalls, M. A., Narendra, D., Noyce, A. J., Ojo, O. O., Okubadejo, N. U., Pagola, A. G., Pastor, P., Errazquin, F. P., Periñán-Tocino, T., Pihlstrom, L., Plun-Favreau, H., Quinn, J., R'Bibo, L., Reed, X., Rezola, E. M., Rizig, M., Rizzu, P., Robak, L., Rodriguez, A. S., Rouleau, G. A., Ruiz-Martínez, J., Ruz, C., Ryten, M., Sadykova, D., Scholz, S. W., Schreglmann, S., Schulte, C., Sharma, M., Shashkin, C., Shulman, J. M., Sierra, M., Siitonen, A., Simón-Sánchez, J., Singleton, A. B., Suarez-Sanmartin, E., Taba, P., Tabernero, C., Tan, M. X., Tartari, J. P., Tejera-Parrado, C., Toft, M., Tolosa, E., Trabzuni, D., Valldeoriola, F., van Hilten, J. J., Keuren-Jensen, K. V., Vargas-González, L., Vela, L., Vives, F., Williams, N., Wood, N. W., Zharkinbekova, N., Zharmukhanov, Z., Zholdybayeva, E., Zimprich, A., Ylikotila, P., Shulman, L. M., von Coelln, R., Reich, S., Savitt, J., Agee, M., Alipanahi, B., Auton, A., Bell, R. K., Bryc, K., Elson, S. L., Fontanillas, P., Furlotte, N. A., Huber, K. E., Hicks, B., Jewett, E. M., Jiang, Y., Kleinman, A., Lin, K.-H., Litterman, N. K., McCreight, J. C., McIntyre, M. H., McManus, K. F., Mountain, J. L., Noblin, E. S., Northover, C. A., Pitts, S. J., Poznik, G. D., Sathirapongsasuti, J. F., Shelton, J. F., Shringarpure, S., Tian, C., Tung, J., Vacic, V., Wang, X., Wilson, C. H., Anderson, T., Bentley, S., Dalrymple-Alford, J., Fowdar, J., Gratten, J., Halliday, G., Henders, A. K., Hickie, I., Kassam, I., Kennedy, M., Kwok, J., Lewis, S., Mellick, G., Montgomery, G., Pearson, J., Pitcher, T., Sidorenko, J., Silburn, P. A., Vallerga, C. L., Visscher, P. M., Wallace, L., Wray, N. R., Xue, A., Yang, J., and Zhang, F. Identification of novel risk loci, causal insights, and heritable risk for parkinson's disease: a meta-analysis of genome-wide association

studies. *The Lancet Neurology 18* (12 2019), 1091–1102. 2.2.9, 2.3.7

[223] Nalls, M. A., Pankratz, N., Lill, C. M., Do, C. B., Hernandez, D. G., Saad, M., DeStefano, A. L., Kara, E., Bras, J., Sharma, M., Schulte, C., Keller, M. F., Arepalli, S., Letson, C., Edsall, C., Stefansson, H., Liu, X., Pliner, H., Lee, J. H., Cheng, R., Ikram, M. A., Ioannidis, J. P. A., Hadjigeorgiou, G. M., Bis, J. C., Martinez, M., Perlmutter, J. S., Goate, A., Marder, K., Fiske, B., Sutherland, M., Xiromerisiou, G., Myers, R. H., Clark, L. N., Stefansson, K., Hardy, J. A., Heutink, P., Chen, H., Wood, N. W., Houlden, H., Payami, H., Brice, A., Scott, W. K., Gasser, T., Bertram, L., Eriksson, N., Foroud, T., and Singleton, A. B. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for parkinson's disease. *Nature Genetics 46* (9 2014), 989–993. 2.1.2

[224] Nelson, M. E., Jester, D. J., Petkus, A. J., and Andel, R. Cognitive reserve, alzheimer's neuropathology, and risk of dementia: A systematic review and meta-analysis. *Neuropsychology Review 31* (6 2021), 233–250. 2.1.3

[225] Nelson, P. T., Schmitt, F. A., Jicha, G. A., Kryscio, R. J., Abner, E. L., Smith, C. D., Eldik, L. J. V., and Markesbery, W. R. Association between male gender and cortical lewy body pathology in large autopsy series. *Journal of Neurology 257* (11 2010), 1875–1881. 2.3.2, 2.4.5

[226] Neumeyer, S., Hemani, G., and Zeggini, E. Strengthening causal inference for complex disease using molecular quantitative trait loci. *Trends in Molecular Medicine 26* (2 2020), 232–241. 5.1.9, 1

[227] Niewiadomska, G., Niewiadomski, W., Steczkowska, M., and Gasiorowska, A. Tau oligomers neurotoxicity. *Life 11* (1 2021), 28. 3.1.2, 3.4

[228] Nishi, A., Matamales, M., Musante, V., Valjent, E., Kuroiwa, M., Kitahara, Y., Rebholz, H., Greengard, P., Girault, J.-A., and Nairn, A. C. Glutamate counteracts dopamine/pka signaling via dephosphorylation of darpp-32 ser-97

and alteration of its cytonuclear distribution. *Journal of Biological Chemistry 292* (1 2017), 1462–1476. 2.3.7

[229] NONPROFIT ORGANIZATION, T. P. F. Statistics, 2023. 2.1

[230] NOVIKOVA, G., KAPOOR, M., TCW, J., ABUD, E. M., EFTHYMIOU, A. G., CHEN, S. X., CHENG, H., FULLARD, J. F., BENDL, J., LIU, Y., ROUSSOS, P., BJÖRKEGREN, J. L., LIU, Y., POON, W. W., HAO, K., MARCORA, E., AND GOATE, A. M. Integration of alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nature Communications 12* (3 2021), 1610. 2.4.4

[231] NULAND, A. J. M., OUDEN, H. E. M., ZACH, H., DIRKX, M. F. M., ASTEN, J. J. A., SCHEENEN, T. W. J., TONI, I., COOLS, R., AND HELMICH, R. C. Gabaergic changes in the thalamocortical circuit in parkinson's disease. *Human Brain Mapping 41* (3 2020), 1017–1029. 2

[232] ODDO, S. The ubiquitin-proteasome system in alzheimer's disease. *Journal of Cellular and Molecular Medicine 12* (4 2008), 363–373. 1

[233] OKSANEN, M., LEHTONEN, S., JARONEN, M., GOLDSTEINS, G., HÄMÄLÄINEN, R. H., AND KOISTINAHO, J. Astrocyte alterations in neurodegenerative pathologies and their modeling in human induced pluripotent stem cell platforms. *Cellular and Molecular Life Sciences 76* (7 2019), 2739–2760. 3.3.6, 3.4.3

[234] ON AGING, N. I. What happens to the brain in alzheimer's disease?, 2022. 1.2.2

[235] ORGANISATION, W. H. Global action plan on the public health response to dementia 2017 - 2025, 2022. 1.2.2

[236] OUELLETTE, J., AND LACOSTE, B. From neurodevelopmental to neurodegenerative disorders: The vascular continuum. *Frontiers in Aging Neuroscience 13* (10 2021). 4.1

[237] OUTEIRO, T. F., KOSS, D. J., ERSKINE, D., WALKER, L., KURZAWA-AKANBI, M., BURN, D., DONAGHY, P., MORRIS, C., TAYLOR, J.-P., THOMAS, A., ATTEMS, J.,

and McKeith, I. Dementia with lewy bodies: an update and outlook. *Molecular Neurodegeneration 14* (12 2019), 5. 2.1.1

[238] Parenti, I., Rabaneda, L. G., Schoen, H., and Novarino, G. Neurodevelopmental disorders: From genetics to functional pathways. *Trends in Neurosciences 43* (8 2020). 1.2.1, 4.1, 4.3.4

[239] Pascal, L. E., True, L. D., Campbell, D. S., Deutsch, E. W., Risk, M., Coleman, I. M., Eichner, L. J., Nelson, P. S., and Liu, A. Y. Correlation of mrna and protein levels: Cell type-specific gene expression of cluster designation antigens in the prostate. *BMC Genomics 9* (2008), 246. 5.1.7

[240] Patel, D., Zhang, X., Farrell, J. J., Chung, J., Stein, T. D., Lunetta, K. L., and Farrer, L. A. Cell-type-specific expression quantitative trait loci associated with alzheimer disease in blood and brain tissue. *Translational Psychiatry 11* (4 2021), 250. 2.4.4

[241] Perl, D. P. Neuropathology of alzheimer's disease. *The Mount Sinai journal of medicine, New York 77* (2010), 32–42. 3.1.1, 3.1.2

[242] Perry, R. H., Irving, D., Blessed, G., Fairbairn, A., and Perry, E. K. Senile dementia of lewy body type. *Journal of the Neurological Sciences 95* (2 1990), 119–139. 2.1

[243] Perucca, E. Pharmacological and therapeutic properties of valproate. *CNS Drugs 16* (2002), 695–714. 4.1.2

[244] Poewe, W., Gauthier, S., Aarsland, D., Leverenz, J. B., Barone, P., Weintraub, D., Tolosa, E., and Dubois, B. Diagnosis and management of parkinson's disease dementia. *International journal of clinical practice 62* (10 2008), 1581–7. 2.1

[245] Poewe, W., Seppi, K., and Tanner, C. Parkinson disease. *Nat Rev Dis Primers 3*, 1. et al. 1.2.2

[246] Porcu, E., Rüeger, S., Lepik, K., Agbessi, M., Ahsan, H., Alves, I., Andiappan, A., Arindrarto, W., Awadalla, P., Battle, A., Beutner, F., Bonder, M. J., Boomsma, D., Christiansen, M., Claringbould, A., Deelen, P., Esko, T., Favé, M.-J., Franke, L., Frayling, T., Gharib, S. A., Gibson, G., Heijmans, B. T., Hemani, G., Jansen, R., Kähönen, M., Kalnapenkis, A., Kasela, S., Kettunen, J., Kim, Y., Kirsten, H., Kovacs, P., Krohn, K., Kronberg-Guzman, J., Kukushkina, V., Lee, B., Lehtimäki, T., Loeffler, M., Marigorta, U. M., Mei, H., Milani, L., Montgomery, G. W., Müller-Nurasyid, M., Nauck, M., Nivard, M., Penninx, B., Perola, M., Pervjakova, N., Pierce, B. L., Powell, J., Prokisch, H., Psaty, B. M., Raitakari, O. T., Ripatti, S., Rotzschke, O., Saha, A., Scholz, M., Schramm, K., Seppälä, I., Slagboom, E. P., Stehouwer, C. D. A., Stumvoll, M., Sullivan, P., 't Hoen, P. A. C., Teumer, A., Thiery, J., Tong, L., Tönjes, A., van Dongen, J., van Iterson, M., van Meurs, J., Veldink, J. H., Verlouw, J., Visscher, P. M., Völker, U., Võsa, U., Westra, H.-J., Wijmenga, C., Yaghootkar, H., Yang, J., Zeng, B., Zhang, F., Arindrarto, W., Beekman, M., Boomsma, D. I., Bot, J., Deelen, J., Deelen, P., Franke, L., Heijmans, B. T., 't Hoen, P. A. C., Hofman, B. A., Hottenga, J. J., Isaacs, A., Bonder, M. J., Jhamai, P. M., Jansen, R., Kielbasa, S. M., Lakenberg, N., Luijk, R., Mei, H., Moed, M., Nooren, I., Pool, R., Schalkwijk, C. G., Slagboom, P. E., Stehouwer, C. D. A., Suchiman, H. E. D., Swertz, M. A., Tigchelaar, E. F., Uitterlinden, A. G., van den Berg, L. H., van der Breggen, R., van der Kallen, C. J. H., van Dijk, F., van Dongen, J., van Duijn, C. M., van Galen, M., van Greevenbroek, M. M. J., van Heemst, D., van Iterson, M., van Meurs, J., van Rooij, J., van't Hof, P., van Zwet, E. W., Vermaat, M., Veldink, J. H., Verbiest, M., Verkerk, M., Wijmenga, C., Zhernakova, D. V., Zhernakova, S., Santoni, F. A., Reymond, A., and Kutalik, Z. Mendelian randomization integrating gwas and eqtl data reveals genetic determinants of complex and clinical traits. *Nature Communications 10* (7 2019), 3300. 1

[247] Porter, R. S., Jaamour, F., and Iwase, S. Neuron-specific alternative splicing of transcriptional machineries: Implications for neurodevelopmental disorders. *Molecular and Cellular Neuroscience 87* (3 2018). 4.4.4

[248] Postuma, R. B., Berg, D., Stern, M., Poewe, W., Olanow, C. W., Oertel, W., Obeso, J., Marek, K., Litvan, I., Lang, A. E., Halliday, G., Goetz, C. G., Gasser, T., Dubois, B., Chan, P., Bloem, B. R., Adler, C. H., and Deuschl, G. Mds clinical diagnostic criteria for parkinson's disease. *Movement Disorders 30* (10 2015), 1591–1601. 2.2.1

[249] Proft, J., Rzhepetskyy, Y., Lazniewska, J., Zhang, F.-X., Cain, S. M., Snutch, T. P., Zamponi, G. W., and Weiss, N. The cacna1h mutation in the gaers model of absence epilepsy enhances t-type ca2+ currents by altering calnexin-dependent trafficking of cav3.2 channels. *Scientific Reports 7* (9 2017), 11513. 1

[250] Pulit, S. L., Stoneman, C., Morris, A. P., Wood, A. R., Glastonbury, C. A., Tyrrell, J., Yengo, L., Ferreira, T., Marouli, E., Ji, Y., Yang, J., Jones, S., Beaumont, R., Croteau-Chonka, D. C., Winkler, T. W., Hattersley, A. T., Loos, R. J. F., Hirschhorn, J. N., Visscher, P. M., Frayling, T. M., Yaghootkar, H., and Lindgren, C. M. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of european ancestry. *Human Molecular Genetics 28* (1 2019). 4.3.5

[251] Qiu, X., Mao, Q., and Tang, Y. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods 14*, 10, 979–982. 1.1.3

[252] Rasmussen, J., Ewing, A. D., Bodea, L., Bodea, G. O., Gearing, M., and Faulkner, G. J. An early proinflammatory transcriptional response to tau pathology is age-specific and foreshadows reduced tau burden. *Brain Pathology 32* (5 2022). 3.1.5

[253] Reinius, B., and Jazin, E. Prenatal sex differences in the human brain. *Molecular Psychiatry 14* (11 2009), 988–989. 4.2.2

[254] Rho, J. M., Donevan, S. D., and Rogawski, M. A. Direct activation of gabaa receptors by barbiturates in cultured rat hippocampal neurons. *The Journal of Physiology 497* (12 1996), 509–522. 2

[255] Ripke, S., Walters, J., O'Donovan, M., and of the Psychiatric Genomics Consortium, S. W. G. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv* (2020). 4.2.5, 4.3.5

[256] Robinson, J. E., and Gradinaru, V. Dopaminergic dysfunction in neurodevelopmental disorders: recent advances and synergistic technologies to aid basic research. *Current Opinion in Neurobiology 48* (2 2018), 17–29. 1.2.1, 2.1.1

[257] Robinson, M. D., McCarthy, D. J., and Smyth, G. K. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics 26* (1 2010). 1.1.3, 4.2.3, 4.3.6

[258] Rogawski, M. A., and Löscher, W. The neurobiology of antiepileptic drugs. *Nature Reviews Neuroscience 5* (7 2004), 553–564. 4.1.2, 1, 2

[259] Romoli, M., Mazzocchetti, P., D'Alonzo, R., Siliquini, S., Rinaldi, V. E., Verrotti, A., Calabresi, P., and Costa, C. Valproic acid and epilepsy: From molecular mechanisms to clinical evidences. *Current Neuropharmacology 17* (9 2019), 926–946. 4.1.2

[260] Rosario, E. R., Chang, L., Head, E. H., Stanczyk, F. Z., and Pike, C. J. Brain levels of sex steroid hormones in men and women during normal aging and in alzheimer's disease. *Neurobiology of Aging 32* (4 2011), 604–613. 5.1.4

[261] Roullet, F. I., Lai, J. K., and Foster, J. A. In utero exposure to valproic acid and autism — a current review of clinical and animal studies. *Neurotoxicology and Teratology 36* (3 2013), 47–56. 4.1.2

[262] Rousso, D., Pearson, C., Gaber, Z., Miquelajauregui, A., Li, S., Portera-Cailliau, C., Morrisey, E., and Novitch, B. Foxp-mediated suppression of n-

cadherin regulates neuroepithelial character and progenitor maintenance in the cns. *Neuron 74* (4 2012). 4.3.6, 4.4.4

[263] Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol 37*, 5, 547–554. 1.1.3

[264] Saez, I., and Vilchez, D. Protein clearance mechanisms and their demise in age-related neurodegenerative diseases. *AIMS Molecular Science 2* (2015), 1–21. 5.1.4

[265] Saghi, M., InanlooRahatloo, K., Alavi, A., Kahrizi, K., and Najmabadi, H. Intellectual disability associated with craniofacial dysmorphism due to polr3b mutation and defect in spliceosomal machinery. *BMC Medical Genomics 15* (12 2022), 89. 4.1.1

[266] Saini, J., Corneo, B., and Miller, J. Nicotinamide ameliorates disease phenotypes in a human ipsc model of age-related macular degeneration. *Cell Stem Cell 20*, 5, 635–647. 1

[267] Salimi, M., Tabasi, F., and Abdolsamadi, M. Disrupted connectivity in the olfactory bulb-entorhinal cortex-dorsal hippocampus circuit is associated with recognition memory deficit in alzheimer's disease model. *Sci Rep 12*, 1. 71. 1.2.2

[268] Sanders, S. J., Schwartz, G. B., and Farh, K. K.-H. Clinical impact of splicing in neurodevelopmental disorders. *Genome Medicine 12* (12 2020), 36. 4.1, 4.1.1

[269] SantaCruz, K., Lewis, J., Spires, T., Paulson, J., Kotilinek, L., Ingelsson, M., Guimaraes, A., DeTure, M., Ramsden, M., McGowan, E., Forster, C., Yue, M., Orne, J., Janus, C., Mariash, A., Kuskowski, M., Hyman, B., Hutton, M., and Ashe, K. H. Tau suppression in a neurodegenerative mouse model improves memory function. *Science 309* (7 2005), 476–481. 3.1.2, 3.1.5

[270] Satija, R., Farrell, J. A., Gennert, D., Schier, A. F., and Regev, A. Spatial reconstruction of single-cell gene expression data. *Nature Biotechnology 33* (5 2015). 1.1.3, 4.2.6

[271] Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., Rubeis, S. D., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., Brand, H., Schwartz, G., Nguyen, R., Guerrero, E. E., Dias, C., Betancur, C., Cook, E. H., Gallagher, L., Gill, M., Sutcliffe, J. S., Thurm, A., Zwick, M. E., Børglum, A. D., State, M. W., Cicek, A. E., Talkowski, M. E., Cutler, D. J., Devlin, B., Sanders, S. J., Roeder, K., Daly, M. J., Buxbaum, J. D., Aleksic, B., Anney, R., Barbosa, M., Bishop, S., Brusco, A., Bybjerg-Grauholm, J., Carracedo, A., Chan, M. C., Chiocchetti, A. G., Chung, B. H., Coon, H., Cuccaro, M. L., Curró, A., Bernardina, B. D., Doan, R., Domenici, E., Dong, S., Fallerini, C., Fernández-Prieto, M., Ferrero, G. B., Freitag, C. M., Fromer, M., Gargus, J. J., Geschwind, D., Giorgio, E., González-Peñas, J., Guter, S., Halpern, D., Hansen-Kiss, E., He, X., Herman, G. E., Hertz-Picciotto, I., Hougaard, D. M., Hultman, C. M., Ionita-Laza, I., Jacob, S., Jamison, J., Jugessur, A., Kaartinen, M., Knudsen, G. P., Kolevzon, A., Kushima, I., Lee, S. L., Lehtimäki, T., Lim, E. T., Lintas, C., Lipkin, W. I., Lopergolo, D., Lopes, F., Ludena, Y., Maciel, P., Magnus, P., Mahjani, B., Maltman, N., Manoach, D. S., Meiri, G., Menashe, I., Miller, J., Minshew, N., Montenegro, E. M., Moreira, D., Morrow, E. M., Mors, O., Mortensen, P. B., Mosconi, M., Muglia, P., Neale, B. M., Nordentoft, M., Ozaki, N., Palotie, A., Parellada, M., Passos-Bueno, M. R., Pericak-Vance, M., Persico, A. M., Pessah, I., Puura, K., Reichenberg, A., Renieri, A., Riberi, E., Robinson, E. B., Samocha, K. E., Sandin, S., Santangelo, S. L., Schellenberg, G., Scherer, S. W., Schlitt, S., Schmidt, R., Schmitt, L., Silva, I. M., Singh, T., Siper, P. M., Smith, M., Soares, G., Stoltenberg, C., Suren, P., Susser, E., Sweeney, J., Szatmari, P., Tang, L., Tassone, F., Teufel, K., Trabetti, E., del Pilar Trelles, M., Walsh, C. A., Weiss, L. A., Werge, T., Werling, D. M., Wigdor, E. M., Wilkinson, E., Willsey, A. J., Yu, T. W., Yu,

M. H., Yuen, R., Zachi, E., Agerbo, E., Als, T. D., Appadurai, V., Bækvad-Hansen, M., Belliveau, R., Buil, A., Carey, C. E., Cerrato, F., Chambert, K., Churchhouse, C., Dalsgaard, S., Demontis, D., Dumont, A., Goldstein, J., Hansen, C. S., Hauberg, M. E., Hollegaard, M. V., Howrigan, D. P., Huang, H., Maller, J., Martin, A. R., Martin, J., Mattheisen, M., Moran, J., Pallesen, J., Palmer, D. S., Pedersen, C. B., Pedersen, M. G., Poterba, T., Poulsen, J. B., Ripke, S., Schork, A. J., Thompson, W. K., Turley, P., and Walters, R. K. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell 180* (2 2020), 568–584.e23. 4.1.1

[272] Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., Young, H., Zabaneh, D., Hägg, S., Hansell, N. K., Karlsson, I. K., Linnarsson, S., Montgomery, G. W., Muñoz-Manchado, A. B., Quinlan, E. B., Schumann, G., Skene, N. G., Webb, B. T., White, T., Arking, D. E., Avramopoulos, D., Bilder, R. M., Bitsios, P., Burdick, K. E., Cannon, T. D., Chiba-Falek, O., Christoforou, A., Cirulli, E. T., Congdon, E., Corvin, A., Davies, G., Deary, I. J., DeRosse, P., Dickinson, D., Djurovic, S., Donohoe, G., Conley, E. D., Eriksson, J. G., Espeseth, T., Freimer, N. A., Giakoumaki, S., Giegling, I., Gill, M., Glahn, D. C., Hariri, A. R., Hatzimanolis, A., Keller, M. C., Knowles, E., Koltai, D., Konte, B., Lahti, J., Hellard, S. L., Lencz, T., Liewald, D. C., London, E., Lundervold, A. J., Malhotra, A. K., Melle, I., Morris, D., Need, A. C., Ollier, W., Palotie, A., Payton, A., Pendleton, N., Poldrack, R. A., Räikkönen, K., Reinvang, I., Roussos, P., Rujescu, D., Sabb, F. W., Scult, M. A., Smeland, O. B., Smyrnis, N., Starr, J. M., Steen, V. M., Stefanis, N. C., Straub, R. E., Sundet, K., Tiemeier, H., Voineskos, A. N., Weinberger, D. R., Widen, E.,

Yu, J., Abecasis, G., Andreassen, O. A., Breen, G., Christiansen, L., Debrabant, B., Dick, D. M., Heinz, A., Hjerling-Leffler, J., Ikram, M. A., Kendler, K. S., Martin, N. G., Medland, S. E., Pedersen, N. L., Plomin, R., Polderman, T. J. C., Ripke, S., van der Sluis, S., Sullivan, P. F., Vrieze, S. I., Wright, M. J., and Posthuma, D. Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics 50* (7 2018). 4.2.5, 4.3.5

[273] Schramm, A., Köster, J., and Marschall, T. Next-generation rna sequencing reveals differential expression of mycn target genes and suggests the mtor pathway as a promising therapy target in mycn- amplified neuroblastoma. *Int J Cancer 132*, 3. 1. 1

[274] Senn, S. M., Kantor, S., Poulton, I. J., Morris, M. J., Sims, N. A., O'Brien, T. J., and Wark, J. D. Adverse effects of valproate on bone: Defining a model to investigate the pathophysiology. *Epilepsia 51* (2 2010). 4.3.1

[275] Sey, N. Y. A., Hu, B., Mah, W., Fauni, H., McAfee, J. C., Rajarajan, P., Brennand, K. J., Akbarian, S., and Won, H. A computational tool (h-magma) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nature Neuroscience 23* (4 2020), 583–593. 3.2.11

[276] Shafiei, S. S., Guerrero-Muñoz, M. J., and Castillo-Carranza, D. L. Tau oligomers: Cytotoxicity, propagation, and mitochondrial damage. *Frontiers in Aging Neuroscience 9* (4 2017). 4

[277] Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., and Waterston, R. H. Dna sequencing at 40: past, present and future. *Nature 550* (10 2017), 345–353. 1.1.1

[278] Shendure, J., and Ji, H. Next-generation dna sequencing. *Nature Biotechnology 26* (10 2008), 1135–1145. 1.1.1

[279] Shi, H., Zhou, Y., Jia, E., Pan, M., Bai, Y., and Ge, Q. Bias in rna-seq library preparation: Current challenges and solutions. *Biomed Res Int 2021*, 1–11. 24. 1.1.3

[280] Shi, H., Zhou, Y., Jia, E., Pan, M., Bai, Y., Ge, Q., et al. Bias in rna-seq library preparation: current challenges and solutions. *BioMed Research International 2021* (2021). 5.1.7

[281] Sidhu, H. S., Srinivas, R., and Sadhotra, A. Evaluate the effects of long-term valproic acid treatment on metabolic profiles in newly diagnosed or untreated female epileptic patients: A prospective study. *Seizure 48* (5 2017). 4.3.4

[282] Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J. D., Bass, N., Bigdeli, T. B., Breen, G., Bromet, E. J., et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature 604*, 7906 (2022), 509–516. 1.1.3

[283] Skene, N., Bryois, J., and Bakken, T. Genetic identification of brain cell types underlying schizophrenia. *Nat Genet 50*, 6, 825–833. 161. 3

[284] Skene, N. G., Bryois, J., Bakken, T. E., Breen, G., Crowley, J. J., Gaspar, H. A., Giusti-Rodriguez, P., Hodge, R. D., Miller, J. A., Muñoz-Manchado, A. B., O'Donovan, M. C., Owen, M. J., Pardiñas, A. F., Ryge, J., Walters, J. T. R., Linnarsson, S., Lein, E. S., Sullivan, P. F., and Hjerling-Leffler, J. Genetic identification of brain cell types underlying schizophrenia. *Nature Genetics 50* (6 2018), 825–833. 2.2.10, 2.3.7

[285] Skene, N. G., and Grant, S. G. N. Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Frontiers in Neuroscience 10* (1 2016). 2.2.6, 5.1.5

[286] Smit, T., Deshayes, N. A. C., Borchelt, D. R., Kamphuis, W., Middeldorp, J., and Hol, E. M. Reactive astrocytes as treatment targets in alzheimer's disease-systematic review of studies using the appsweps1de9 mouse model. *Glia 69* (8 2021), 1852–1881. 2

[287] Smith, C., Malek, N., Grosset, K., Cullen, B., Gentleman, S., and Grosset, D. G. Neuropathology of dementia in patients with parkinson's disease: a systematic

review of autopsy studies. *Journal of Neurology, Neurosurgery Psychiatry* (8 2019), jnnp–2019–321111. 2.1.1

[288] Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical applications in genetics and molecular biology 3*, 1 (2004). 1.1.3

[289] Snow, W. M., Hartle, K., and Ivanco, T. L. Altered morphology of motor cortex neurons in the vpa rat model of autism. *Developmental Psychobiology 50* (11 2008), 633–639. 4.1.2

[290] Soneson, C., and Delorenzi, M. A comparison of methods for differential expression analysis of rna-seq data. *BMC Bioinformatics 14*, 1. 34. 1.1.3, 2.4.5

[291] Soneson, C., and Robinson, M. D. Bias, robustness and scalability in single-cell differential expression analysis. *Nature Methods 15* (4 2018), 255–261. 1

[292] Spillantini, M. G., Schmidt, M. L., Lee, V. M.-Y., Trojanowski, J. Q., Jakes, R., and Goedert, M. -synuclein in lewy bodies. *Nature 388* (8 1997), 839–840. 2.1.1

[293] Stahl, E. A., Breen, G., Forstner, A. J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J. R. I., Gaspar, H. A., de Leeuw, C. A., Steinberg, S., Pavlides, J. M. W., Trzaskowski, M., Byrne, E. M., Pers, T. H., Holmans, P. A., Richards, A. L., Abbott, L., Agerbo, E., Akil, H., Albani, D., Alliey-Rodriguez, N., Als, T. D., Anjorin, A., Antilla, V., Awasthi, S., Badner, J. A., Bækvad-Hansen, M., Barchas, J. D., Bass, N., Bauer, M., Belliveau, R., Bergen, S. E., Pedersen, C. B., Bøen, E., Boks, M. P., Boocock, J., Budde, M., Bunney, W., Burmeister, M., Bybjerg-Grauholm, J., Byerley, W., Casas, M., Cerrato, F., Cervantes, P., Chambert, K., Charney, A. W., Chen, D., Churchhouse, C., Clarke, T.-K., Coryell, W., Craig, D. W., Cruceanu, C., Curtis, D., Czerski, P. M., Dale, A. M., de Jong, S., Degenhardt, F., Del-Favero, J., DePaulo, J. R., Djurovic, S., Dobbyn, A. L., Dumont, A., Elvsåshagen, T., Escott-Price,

V., Fan, C. C., Fischer, S. B., Flickinger, M., Foroud, T. M., Forty, L., Frank, J., Fraser, C., Freimer, N. B., Frisén, L., Gade, K., Gage, D., Garnham, J., Giambartolomei, C., Pedersen, M. G., Goldstein, J., Gordon, S. D., Gordon-Smith, K., Green, E. K., Green, M. J., Greenwood, T. A., Grove, J., Guan, W., Guzman-Parra, J., Hamshere, M. L., Hautzinger, M., Heilbronner, U., Herms, S., Hipolito, M., Hoffmann, P., Holland, D., Huckins, L., Jamain, S., Johnson, J. S., Juréus, A., Kandaswamy, R., Karlsson, R., Kennedy, J. L., Kittel-Schneider, S., Knowles, J. A., Kogevinas, M., Koller, A. C., Kupka, R., Lavebratt, C., Lawrence, J., Lawson, W. B., Leber, M., Lee, P. H., Levy, S. E., Li, J. Z., Liu, C., Lucae, S., Maaser, A., MacIntyre, D. J., Mahon, P. B., Maier, W., Martinsson, L., McCarroll, S., McGuffin, P., McInnis, M. G., McKay, J. D., Medeiros, H., Medland, S. E., Meng, F., Milani, L., Montgomery, G. W., Morris, D. W., Mühleisen, T. W., Mullins, N., Nguyen, H., Nievergelt, C. M., Adolfsson, A. N., Nwulia, E. A., O'Donovan, C., Loohuis, L. M. O., Ori, A. P. S., Oruc, L., Ösby, U., Perlis, R. H., Perry, A., Pfennig, A., Potash, J. B., Purcell, S. M., Regeer, E. J., Reif, A., Reinbold, C. S., Rice, J. P., Rivas, F., Rivera, M., Roussos, P., Ruderfer, D. M., Ryu, E., Sánchez-Mora, C., Schatzberg, A. F., Scheftner, W. A., Schork, N. J., Weickert, C. S., Shehktman, T., Shilling, P. D., Sigurdsson, E., Slaney, C., Smeland, O. B., Sobell, J. L., Hansen, C. S., Spijker, A. T., Clair, D. S., Steffens, M., Strauss, J. S., Streit, F., Strohmaier, J., Szelinger, S., Thompson, R. C., Thorgeirsson, T. E., Treutlein, J., Vedder, H., Wang, W., Watson, S. J., Weickert, T. W., Witt, S. H., Xi, S., Xu, W., Young, A. H., Zandi, P., Zhang, P., Zöllner, S., Adolfsson, R., Agartz, I., Alda, M., Backlund, L., Baune, B. T., Bellivier, F., Berrettini, W. H., Biernacka, J. M., Blackwood, D. H. R., Boehnke, M., Børglum, A. D., Corvin, A., Craddock, N., Daly, M. J., Dannlowski, U., Esko, T., Etain, B., Frye, M., Fullerton, J. M., Gershon, E. S., Gill, M., Goes, F., Grigoroiu-Serbanescu, M., Hauser, J., Hougaard, D. M., Hult-

man, C. M., Jones, I., Jones, L. A., Kahn, R. S., Kirov, G., Landén, M., Leboyer, M., Lewis, C. M., Li, Q. S., Lissowska, J., Martin, N. G., Mayoral, F., McElroy, S. L., McIntosh, A. M., McMahon, F. J., Melle, I., Metspalu, A., Mitchell, P. B., Morken, G., Mors, O., Mortensen, P. B., Müller-Myhsok, B., Myers, R. M., Neale, B. M., Nimgaonkar, V., Nordentoft, M., Nöthen, M. M., O'Donovan, M. C., Oedegaard, K. J., Owen, M. J., Paciga, S. A., Pato, C., Pato, M. T., Posthuma, D., Ramos-Quiroga, J. A., Ribasés, M., Rietschel, M., Rouleau, G. A., Schalling, M., Schofield, P. R., Schulze, T. G., Serretti, A., Smoller, J. W., Stefansson, H., Stefansson, K., Stordal, E., Sullivan, P. F., Turecki, G., Vaaler, A. E., Vieta, E., Vincent, J. B., Werge, T., Nurnberger, J. I., Wray, N. R., Florio, A. D., Edenberg, H. J., Cichon, S., Ophoff, R. A., Scott, L. J., Andreassen, O. A., Kelsoe, J., and Sklar, P. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature Genetics 51* (5 2019). 4.2.5, 4.3.5

[294] Stamouli, E., and Politis, A. Pro-inflammatory cytokines in alzheimer's disease. *Psychiatriki 27* (10 2016), 264–275. 3

[295] Stamouli, S., Anderlid, B.-M., Willfors, C., Thiruvahindrapuram, B., Wei, J., Berggren, S., Nordgren, A., Scherer, S. W., Lichtenstein, P., Tammimies, K., and Bölte, S. Copy number variation analysis of 100 twin pairs enriched for neurodevelopmental disorders. *Twin Research and Human Genetics 21* (2 2018), 1–11. 4.1.1

[296] Stankovic, I. N., and Colak, D. Prenatal drugs and their effects on the developing brain: Insights from three-dimensional human organoids. *Frontiers in Neuroscience 16* (3 2022). 4.1.2

[297] Stefanis, L. -synuclein in parkinson's disease. *Cold Spring Harbor Perspectives in Medicine 2* (2 2012), a009399–a009399. 2.1.1, 2.1.2, 2.4.1

[298] Stessman, H. A. F., Xiong, B., Coe, B. P., Wang, T., Hoekzema, K., Fenckova, M., Kvarnung, M., Gerdts, J., Trinh, S., Cosemans, N., Vives, L., Lin,

J., Turner, T. N., Santen, G., Ruivenkamp, C., Kriek, M., van Haeringen, A., Aten, E., Friend, K., Liebelt, J., Barnett, C., Haan, E., Shaw, M., Gecz, J., Anderlid, B.-M., Nordgren, A., Lindstrand, A., Schwartz, C., Kooy, R. F., Vandeweyer, G., Helsmoortel, C., Romano, C., Alberti, A., Vinci, M., Avola, E., Giusto, S., Courchesne, E., Pramparo, T., Pierce, K., Nalabolu, S., Amaral, D. G., Scheffer, I. E., Delatycki, M. B., Lockhart, P. J., Hormozdiari, F., Harich, B., Castells-Nobau, A., Xia, K., Peeters, H., Nordenskjöld, M., Schenck, A., Bernier, R. A., and Eichler, E. E. Targeted sequencing identifies 91 neurodevelopmental-disorder risk genes with autism and developmental-disability biases. *Nature Genetics 49* (4 2017), 515–526. 4.1.1

[299] Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. Comprehensive integration of single-cell data. *Cell 177* (6 2019), 1888–1902.e21. 1.1.3

[300] Su, C.-H., D, D., and Tarn, W.-Y. Alternative splicing in neurogenesis and brain development. *Frontiers in Molecular Biosciences 5* (2 2018). 4.4.4

[301] Subramanian, A., Tamayo, P., and Mootha, V. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences 102*, 43, 15545–15550. 42. 1.1.3

[302] Sun, S., Zhu, J., Ma, Y., and Zhou, X. Accuracy, robustness and scalability of dimensionality reduction methods for single-cell rna-seq analysis. *Genome Biology 20* (12 2019), 269. 1.1.3

[303] Takemon, Y., Chick, J. M., Gyuricza, I. G., Skelly, D. A., Devuyst, O., Gygi, S. P., Churchill, G. A., and Korstanje, R. Proteomic and transcriptomic profiling reveal different aspects of aging in the kidney. *eLife 10* (3 2021). 5.1.7

[304] Tan, M. M., Lawton, M. A., Jabbari, E., Reynolds, R. H., Iwaki, H., Blauwendraat, C., Kanavou, S., Pollard, M. I., Hubbard, L., Malek, N., Grosset, K. A., Marrinan, S. L., Bajaj, N., Barker, R. A., Burn, D. J.,

Bresner, C., Foltynie, T., Wood, N. W., Williams-Gray, C. H., Hardy, J., Nalls, M. A., Singleton, A. B., Williams, N. M., Ben-Shlomo, Y., Hu, M. T., Grosset, D. G., Shoai, M., and Morris, H. R. Genome-wide association studies of cognitive and motor progression in parkinson's disease. *Movement Disorders 36* (2 2021), 424–433. 2.1.2

[305] Tanaka, M., Olsen, R. W., Medina, M. T., Schwartz, E., Alonso, M. E., Duron, R. M., Castro-Ortega, R., Martinez-Juarez, I. E., Pascual-Castroviejo, I., Machado-Salas, J., Silva, R., Bailey, J. N., Bai, D., Ochoa, A., Jara-Prado, A., Pineda, G., Macdonald, R. L., and Delgado-Escueta, A. V. Hyperglycosylation and reduced gaba currents of mutated gabrb3 polypeptide in remitting childhood absence epilepsy. *The American Journal of Human Genetics 82* (6 2008), 1249–1261. 4.4.3

[306] Tang, X., Jaenisch, R., and Sur, M. The role of gabaergic signalling in neurodevelopmental disorders. *Nature Reviews Neuroscience 22* (5 2021), 290–307. 2

[307] Tanner, C. M., Kamel, F., Ross, G. W., Hoppin, J. A., Goldman, S. M., Korell, M., Marras, C., Bhudhikanok, G. S., Kasten, M., Chade, A. R., Comyns, K., Richards, M. B., Meng, C., Priestley, B., Fernandez, H. H., Cambi, F., Umbach, D. M., Blair, A., Sandler, D. P., and Langston, J. W. Rotenone, paraquat, and parkinson's disease. *Environmental Health Perspectives 119* (6 2011), 866–872. 2.1.3

[308] Thaler, A., Ash, E., Gan-Or, Z., Orr-Urtreger, A., and Giladi, N. The lrrk2 g2019s mutation as the cause of parkinson's disease in ashkenazi jews. *Journal of Neural Transmission 116* (11 2009), 1473–1482. 2.1.2

[309] Thotala, D., Karvas, R. M., Engelbach, J. A., Garbow, J. R., Hallahan, A. N., DeWees, T. A., Laszlo, A., and Hallahan, D. E. Valproic acid enhances the efficacy of radiation therapy by protecting normal hippocampal neurons and sensitizing malignant glioblastoma cells. *Oncotarget 6* (10 2015), 35004–35022. 4.1.2

[310] Tilley, B. S. A clinico-pathological investigation of the human basal ganglia and brainstem in clinical subtypes of parkinson's disease. 2.1.3, 2.2.1

[311] Timp, W., and Timp, G. Beyond mass spectrometry, the next step in proteomics. *Science Advances 6* (1 2020). 5.1.7

[312] Tong, V., Teng, X. W., Chang, T. K. H., and Abbott, F. S. Valproic acid ii: Effects on oxidative stress, mitochondrial membrane potential, and cytotoxicity in glutathione-depleted rat hepatocytes. *Toxicological Sciences 86* (8 2005). 4

[313] Traag, V., Waltman, L., and Eck, N. From louvain to leiden: guaranteeing well-connected communities. *Sci Rep 9*, 1. 1.1.3

[314] Trabzuni, D., Ramasamy, A., Imran, S., Walker, R., Smith, C., Weale, M. E., Hardy, J., and Ryten, M. Widespread sex differences in gene expression and splicing in the adult human brain. *Nature Communications 4* (12 2013). 5.1.4

[315] Tracy, T. E., and Gan, L. Tau-mediated synaptic and neuronal dysfunction in neurodegenerative disease. *Current Opinion in Neurobiology 51* (8 2018), 134–138. 1, 1

[316] Trapnell, C., Cacchiarelli, D., and Grimsby, J. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol 32*, 4, 381–386. 51. 1.1.2, 1.1.3

[317] Tredici, K. D., and Braak, H. Idiopathic parkinson's disease: Staging an -synucleinopathy with a predictable pathoanatomy. *Madame Curie Bioscience Database* (2000). 2.1.1

[318] (US), N. A. P. Clinical characteristics of intellectual disabilities. 1.2.1

[319] V., G. R., JL, Y.-G., and MJ, M. Rna-sequencing from single nuclei. In *Proceedings of the National Academy of Sciences. Published online 2013.* 19. 1.1.1

[320] Vajda, F. J., O'Brien, T. J., Graham, J. E., Lander, C. M., and Eadie, M. J. Dose dependence of fetal malformations associated with valproate. *Neurology 81* (9 2013), 999–1003. 4.1.2

[321] Velez-Pardo, C., Lorenzo-Betancor, O., and Jimenez-Del-Rio, M. The distribution and risk effect of gba variants in a large cohort of pd patients from colombia and peru. *Parkinsonism Relat Disord 63*, 204–208. 118. 2.1.2, 2.4.1

[322] Vieth, B., Parekh, S., Ziegenhain, C., Enard, W., and Hellmann, I. A systematic evaluation of single cell rna-seq analysis pipelines. *Nature Communications 10* (10 2019), 4667. 1.1.2

[323] Vilchez, D., Saez, I., and Dillin, A. The role of protein clearance mechanisms in organismal ageing and age-related diseases. *Nature Communications 5* (12 2014), 5659. 5.1.4

[324] Visscher, P., Brown, M., McCarthy, M., and Yang, J. Five years of gwas discovery. *The American Journal of Human Genetics 90*, 1, 7–24. 45. 1.1.3

[325] Vissers, L., Gilissen, C., and Veltman, J. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet 17*, 1, 9–18. 1.2.1

[326] Vogel, C., and Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics 13* (4 2012), 227–232. 5.1.7

[327] Vogt, B. Cingulate cortex in parkinson's disease. *In 2019*, 253–266. 2.1.4

[328] Volpicelli-Daley, L., and Brundin, P. Prion-like propagation of pathology in parkinson disease. *In 2018*, 321–335. 2.1.1

[329] Vom Berg, J., Prokop, S., Miller, K. R., Obst, J., Kälin, R. E., Lopategui-Cabezas, I., Wegner, A., Mair, F., Schipke, C. G., Peters, O., et al. Inhibition of il-12/il-23 signaling reduces alzheimer's disease–like pathology and cognitive decline. *Nature medicine 18*, 12 (2012), 1812–1819. 4

[330] Vossler, D. G. Comparative risk of major congenital malformations with 8 different antiepileptic drugs: A prospective cohort study of the eurap registry. *Epilepsy Currents 19* (3 2019). 4.1.2, 4.4.5

[331] Walls, A. B., Nilsen, L. H., Eyjolfsson, E. M., Vestergaard, H. T., Hansen, S. L., Schousboe, A., Sonnewald, U., and Waagepetersen, H. S. Gad65 is essential for synthesis of gaba destined for tonic inhibition regulating epileptiform activity. *Journal of Neurochemistry 115* (12 2010). 4.3.6, 4.4.4

[332] Wang, D., Liu, S., Warrell, J., Won, H., Shi, X., Navarro, F. C. P., Clarke, D., Gu, M., Emani, P., Yang, Y. T., Xu, M., Gandal, M. J., Lou, S., Zhang, J., Park, J. J., Yan, C., Rhie, S. K., Manakongtreecheep, K., Zhou, H., Nathan, A., Peters, M., Mattei, E., Fitzgerald, D., Brunetti, T., Moore, J., Jiang, Y., Girdhar, K., Hoffman, G. E., Kalayci, S., Gümüş, Z. H., Crawford, G. E., Roussos, P., Akbarian, S., Jaffe, A. E., White, K. P., Weng, Z., Sestan, N., Geschwind, D. H., Knowles, J. A., Gerstein, M. B., Ashley-Koch, A. E., Crawford, G. E., Garrett, M. E., Song, L., Safi, A., Johnson, G. D., Wray, G. A., Reddy, T. E., Goes, F. S., Zandi, P., Bryois, J., Jaffe, A. E., Price, A. J., Ivanov, N. A., Collado-Torres, L., Hyde, T. M., Burke, E. E., Kleiman, J. E., Tao, R., Shin, J. H., Akbarian, S., Girdhar, K., Jiang, Y., Kundakovic, M., Brown, L., Kassim, B. S., Park, R. B., Wiseman, J. R., Zharovsky, E., Jacobov, R., Devillers, O., Flatow, E., Hoffman, G. E., Lipska, B. K., Lewis, D. A., Haroutunian, V., Hahn, C.-G., Charney, A. W., Dracheva, S., Kozlenkov, A., Belmont, J., DelValle, D., Francoeur, N., Hadjimichael, E., Pinto, D., van Bakel, H., Roussos, P., Fullard, J. F., Bendl, J., Hauberg, M. E., Mangravite, L. M., Peters, M. A., Chae, Y., Peng, J., Niu, M., Wang, X., Webster, M. J., Beach, T. G., Chen, C., Jiang, Y., Dai, R., Shieh, A. W., Liu, C., Grennan, K. S., Xia, Y., Vadukapuram, R., Wang, Y., Fitzgerald, D., Cheng, L., Brown, M., Brown, M., Brunetti, T., Goodman, T., Alsayed, M., Gandal, M. J., Geschwind, D. H., Won, H., Polioudakis, D., Wamsley, B., Yin, J., Hadzic, T., Ubieta, L. D. L. T., Swarup, V., Sanders, S. J., State, M. W., Werling, D. M., An, J.-Y., Sheppard, B., Willsey, A. J., White, K. P., Ray, M., Giase, G., Kefi, A., Mattei, E., Purcaro, M., Weng, Z., Moore, J., Pratt, H., Huey, J., Borrman, T., Sullivan,

P. F., Giusti-Rodriguez, P., Kim, Y., Sullivan, P., Szatkiewicz, J., Rhie, S. K., Armoskus, C., Camarena, A., Farnham, P. J., Spitsyna, V. N., Witt, H., Schreiner, S., Evgrafov, O. V., Knowles, J. A., Gerstein, M., Liu, S., Wang, D., Navarro, F. C. P., Warrell, J., Clarke, D., Emani, P. S., Gu, M., Shi, X., Xu, M., Yang, Y. T., Kitchen, R. R., Gürsoy, G., Zhang, J., Carlyle, B. C., Nairn, A. C., Li, M., Pochareddy, S., Sestan, N., Skarica, M., Li, Z., Sousa, A. M. M., Santpere, G., Choi, J., Zhu, Y., Gao, T., Miller, D. J., Cherskov, A., Yang, M., Amiri, A., Coppola, G., Mariani, J., Scuderi, S., Szekely, A., Vaccarino, F. M., Wu, F., Weissman, S., Roychowdhury, T., and Abyzov, A. Comprehensive functional genomic resource and integrative model for the human brain. *Science 362* (12 2018). 2.2.5, 2.3.3, 3.2.8, 3.3.2

[333] Wang, T., Li, B., Nelson, C. E., and Nabavi, S. Comparative analysis of differential gene expression analysis tools for single-cell rna sequencing data. *BMC Bioinformatics 20* (12 2019), 40. 1.1.3, 2.4.5, 1

[334] WANG, Y., LIU, J., HUANG, B., XU, Y.-M., LI, J., HUANG, L.-F., LIN, J., ZHANG, J., MIN, Q.-H., YANG, W.-M., and WANG, X.-Z. Mechanism of alternative splicing and its regulation. *Biomedical Reports 3* (3 2015). 4.4.4

[335] Wang, Z., Chan, A. Y. L., Coghill, D., Ip, P., Lau, W. C. Y., Simonoff, E., Brauer, R., Wei, L., Wong, I. C. K., and Man, K. K. C. Association between prenatal exposure to antipsychotics and attention-deficit/hyperactivity disorder, autism spectrum disorder, preterm birth, and small for gestational age. *JAMA Internal Medicine 181* (10 2021). 4.1.2, 4.4.5

[336] Wei, Z., Zeng, K., Hu, J., Li, X., Huang, F., Zhang, B., Wang, J.-Z., Liu, R., Li, H.-L., and Wang, X. Usp10 deubiquitinates tau, mediating its aggregation. *Cell Death Disease 13* (8 2022), 726. 2

[337] Weiss, N., and Zamponi, G. W. Genetic t-type calcium channelopathies. *Journal of Medical Genetics 57* (1 2020), 1–10. 1

[338] Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., and Macosko, E. Z. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell 177* (6 2019), 1873–1887.e17. 1.1.3

[339] Wen, X., Pique-Regi, R., and Luca, F. Integrating molecular qtl data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genetics 13* (3 2017), e1006646. 2

[340] WENDELL, A. D. Overview and epidemiology of substance abuse in pregnancy. *Clinical Obstetrics Gynecology 56* (3 2013), 91–96. 4.1.2

[341] Wilfert, A. B., Sulovari, A., Turner, T. N., Coe, B. P., and Eichler, E. E. Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications. *Genome Medicine 9* (12 2017), 101. 4.1.1

[342] Witzig, V. S., Komnig, D., and Falkenburger, B. H. Changes in striatal medium spiny neuron morphology resulting from dopamine depletion are reversible. *Cells 9* (11 2020), 2441. 3

[343] Wyss-Coray, T. Ageing, neurodegeneration and brain rejuvenation. *Nature 539*, 7628, 180–186. 1.2.2, 5.1.4

[344] Würtz, A., Rytter, D., Vestergaard, C., Christensen, J., Vestergaard, M., and Bech, B. Prenatal exposure to antiepileptic drugs and use of primary healthcare during childhood: a population-based cohort study in denmark. *BMJ Open 7*, 1. 233. 4.1.2

[345] Xia, C., Makaretz, S. J., Caso, C., McGinnis, S., Gomperts, S. N., Sepulcre, J., Gomez-Isla, T., Hyman, B. T., Schultz, A., Vasdev, N., Johnson, K. A., and Dickerson, B. C. Association of in vivo [¡sup¿18¡/sup¿ f]av-1451 tau pet imaging results with cortical atrophy and symptoms in typical and atypical alzheimer disease. *JAMA Neurology 74* (4 2017), 427. 3.1.2

[346] Xu, Y., Yang, J., and Shang, H. Meta-analysis of risk factors for parkinson's disease dementia. *Translational Neurodegeneration 5* (12 2016), 11. 2.1.3

[347] Xu, Y., Zhao, M., Han, Y., and Zhang, H. Gabaergic inhibitory interneuron deficits in alzheimer's disease: Implications for treatment. *Frontiers in Neuroscience 14* (6 2020). 2

[348] Yadav, S. The wholeness in suffix -omics, -omes, and the word om. *J Biomol Tech 18*, 5. 1.1

[349] Yang, W., Liu, J., Zheng, F., Jia, M., Zhao, L., Lu, T., Ruan, Y., Zhang, J., Yue, W., Zhang, D., and Wang, L. The evidence for association of atp2b2 polymorphisms with autism in chinese han population. *PLoS ONE 8* (4 2013), e61021. 4.4.3

[350] Yeung, E. S. Genome-wide correlation between mrna and protein in a single cell. *Angewandte Chemie International Edition 50* (1 2011), 583–585. 5.1.7

[351] Yoshiyama, Y., Higuchi, M., Zhang, B., Huang, S.-M., Iwata, N., Saido, T., Maeda, J., Suhara, T., Trojanowski, J. Q., and Lee, V. M.-Y. Synapse loss and microglial activation precede tangles in a p301s tauopathy mouse model. *Neuron 53* (2 2007), 337–351. 3.1.2

[352] Yu, G. Gene ontology semantic similarity analysis using gosemsim. *In 2020*, 207–215. 1.1.2, 2.2.8

[353] Zaghlool, A., Niazi, A., Åsa K. Björklund, Westholm, J. O., Ameur, A., and Feuk, L. Characterization of the nuclear and cytosolic transcriptomes in human brain tissue reveals new insights into the subcellular distribution of rna transcripts. *Scientific Reports 11* (2 2021), 4076. 5

[354] Zaja-Milatovic, S., Keene, C., Montine, K., Leverenz, J., Tsuang, D., and Montine, T. Selective dendritic degeneration of medium spiny neurons in dementia with lewy bodies. *Neurology 66*, 10, 1591–1593. 162. 2.3.7, 2.4.2, 3

[355] Zech, M., Lam, D. D., Weber, S., Berutti, R., Poláková, K., Havránková, P., Fečíková, A., Strom, T. M., Růžička, E., Jech, R., and Winkelmann, J.

A unique de novo gain-of-function variant in camk4 associated with intellectual disability and hyperkinetic movement disorder. *Cold Spring Harbor molecular case studies 4* (12 2018). 4.4.3

[356] Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., Häring, M., Braun, E., Borm, L. E., Manno, G. L., Codeluppi, S., Furlan, A., Lee, K., Skene, N., Harris, K. D., Hjerling-Leffler, J., Arenas, E., Ernfors, P., Marklund, U., and Linnarsson, S. Molecular architecture of the mouse nervous system. *Cell 174* (8 2018). 6

[357] Zeng, L., Yang, J., Peng, S., Zhu, J., Zhang, B., Suh, Y., and Tu, Z. Transcriptome analysis reveals the difference between "healthy" and "common" aging and their connection with age-related diseases. *Aging Cell 19* (3 2020). 5.1.4

[358] Zeng, P., Shao, Z., and Zhou, X. Statistical methods for mediation analysis in the era of high-throughput genomics: Current successes and future challenges. *Computational and Structural Biotechnology Journal 19* (2021), 3209–3224. 4

[359] Zeng, W., Jiang, S., Kong, X., El-Ali, N., Ball, A. R., Ma, C. I.-H., Hashimoto, N., Yokomori, K., and Mortazavi, A. Single-nucleus rna-seq of differentiating human myoblasts reveals the extent of fate heterogeneity. *Nucleic Acids Research* (8 2016), gkw739. 5.1.6

[360] Zhang, R., Manza, P., and Volkow, N. D. Prenatal caffeine exposure: association with neurodevelopmental outcomes in 9- to 11-year-old children. *Journal of Child Psychology and Psychiatry 63* (5 2022), 563–578. 4.1.2

[361] Zhao, H., Wang, Q., Yan, T., Zhang, Y., juan Xu, H., peng Yu, H., Tu, Z., Guo, X., hui Jiang, Y., jiang Li, X., Zhou, H., and Zhang, Y. Q. Maternal valproic acid exposure leads to neurogenesis defects and autism-like behaviors in non-human primates. *Translational Psychiatry 9* (12 2019). 4.4.3

[362] Zheng, C., Zhou, X.-W., and Wang, J.-Z. The dual roles of cytokines in alzheimer's disease: update on interleukins, tnf-, tgf- and ifn-. *Translational Neurodegeneration 5* (12 2016), 7. 3

[363] Zheng, G., Terry, J., and Belgrader, P. Massively parallel digital transcriptional profiling of single cells. *Nat Commun 8*, 1. 22. 1.1.3

[364] Zhou, F.-M. The striatal medium spiny neurons: what they are and how they link with parkinson's disease. In *Genetics, Neurology, Behavior, and Diet in Parkinson's Disease*. Elsevier, 2020, pp. 395–412. 3

[365] Zhou, J., Gennatas, E., Kramer, J., Miller, B., and Seeley, W. Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron 73* (3 2012), 1216–1227. 3.1.2

[366] Zhu, Y., Wang, L., Yin, Y., and Yang, E. Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Scientific Reports 7* (7 2017), 5435. 5.1.4

[367] Zimmerman, K. D., Evans, C., and Langefeld, C. D. Reply to: A balanced measure shows superior performance of pseudobulk methods in single-cell rna-sequencing analysis. *Nature Communications 13* (12 2022), 7852. 5.1.8

[368] Éva Mészáros. Dna sequencing methods: from sanger to ngs. 1.1.1

[369] 'Fleming S, 'Marioni J, M. Cellbender remove-background: a deep generative model for unsupervised removal of background noise from scrna-seq datasets. 23. 1.1.3, 6