



Queen Mary
University of London

Integrative Omics Data Analysis to Identify Shared Genes and Pathways Across Neurodegenerative Diseases

Khadija Paderwala

August 2023

This research dissertation is submitted for the MSc in Bioinformatics
at Queen Mary,
University of London

Centre for Neuroscience, Surgery, and Trauma, Blizard Institute
Barts and the London School of Medicine and Dentistry
Queen Mary University of London

Supervisor: Dr Sarah Louise Morgan

Word count (excluding tables and captions): 10,016

Abstract

Background: Between 2015 and 2050 the proportion of the world's population over 60 will nearly double from 12% to 22%. With an aging population comes the increased risk of neurodegenerative diseases, a group of disorders effecting neurons in the brain resulting in their progressive atrophy. Despite years of research, a cure is yet to be found, highlighting the importance of continued research of therapeutic options. This can only be achieved by understanding the implicated genes and pathways that can be targeted with treatment. Many neurodegenerative diseases share similarities in risk factors, genetic determinants, symptoms, pathology and disease mechanisms, prompting the hypothesis that these diseases are different manifestations of the same underlying disease and treatment for one may be able to treat others.

Aims: This study aimed to identify overlapping dysfunctional genes and pathways across the following diseases to determine potential drug targets: Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis, Lewy body dementia, Huntington's disease and multiple sclerosis.

Methodology: Data was gathered from the most comprehensive genome-wide association study for each disease and filtered to retain only significant genetic variants. These were then mapped to genes and pathways using bioinformatic tools. The overlapping genes and pathways were identified and reported.

Key findings: Analysis revealed overlapping genes across all the diseases when compared pairwise apart from HD. There were 8 genes overlapping across 3 or more diseases. The following genes were observed across AD, PD, ALS and MS: BTNL2, HLA-DQA1, HLA-DQA2, HLA-DRA. HLA-DRB5 was exclusive to AD, ALS and MS while HLA-DRB1 was exclusive to AD, PD and MS. DGKQ and TMEM175 were identified across PD, ALS and LBD. The top pathways enriched in four diseases related to the MHC class II, antigen processing and presentation, T-cell differentiation and activation, immunoglobulin-mediated immune response, the synaptic vesicle cycle, endocytosis, regulation of monooxygenase activity and somatic cell DNA recombination.

Acknowledgements

I am deeply grateful to my supervisor Dr Sarah Morgan for their invaluable guidance and patient assistance throughout the journey of completing this thesis. When I faced challenges or felt stuck, your insightful advice helped me navigate the complexities of my project. I am entirely thankful for your mentorship and dedication to my academic growth. Your insightful suggestions and thoughtful feedback not only improved the quality of my project but enriched by understating, enabling me to discover my true passion for this field of word. I cannot put into words the privilege it has been to have you as my supervisor.

I would also like to express my heartfelt appreciation to my boyfriend Daniel Camarinha, for his unwavering support and encouragement during this demanding phase of my academic pursuit. Your belief in my abilities, patient listening and full faith in my potential gave me the strength to overcome challenges and self-doubt. Thank you for celebrating every academic milestone with me and pushing me to pursuit excellence. Your presence has been a constant source of comfort, solace and positivity for me.

Lastly, I'd like to express my thanks to the new lifelong friends I made this year. You all played a pivotal role during this MSc. The way we helped each other during other projects and tackled difficult tasks as a group, not only enhanced the quality of our work, but made the journey more enjoyable.

Table of Contents

Abstract.....	2
Acknowledgements	3
List of figures.....	6
List of tables.....	8
Abbreviations	9
1 Introduction.....	11
<i>1.1 Neurodegenerative diseases.....</i>	<i>11</i>
1.1.1 Alzheimer's disease	11
1.1.2 Parkinson's disease	12
1.1.3 Amyotrophic lateral sclerosis	13
1.1.4 Lewy body dementia.....	13
1.1.5 Huntington's disease	14
1.1.6 Multiple sclerosis	15
<i>1.2 Genetics.....</i>	<i>16</i>
1.2.1 Genetics of Alzheimer's disease	17
1.2.2 Genetics of Parkinson's disease	18
1.2.3 Genetics of amyotrophic lateral sclerosis	19
1.2.4 Genetics of Lewy body dementia	20
1.2.5 Genetics of Huntington's disease.....	21
1.2.6 Genetics of multiple sclerosis	22
<i>1.3 Commonalities across neurodegenerative diseases.....</i>	<i>23</i>
1.3.1 Shared risk factors.....	23
1.3.2 Shared symptoms	24
1.3.3 Shared pathology	25
1.3.4 Shared disease mechanisms	26
<i>1.4 Justification for study, hypothesis and aims</i>	<i>28</i>
2 Methods.....	30
<i>2.1 Input data</i>	<i>30</i>
<i>2.2 Analysis</i>	<i>30</i>
<i>2.3 Data cleaning and filtering.....</i>	<i>32</i>
<i>2.4 GREAT analysis</i>	<i>33</i>
2.4.1 Gene overlap	33
<i>2.5 Pathway enrichment.....</i>	<i>34</i>
2.5.1 Pathway overlap.....	34
3 Results	36
<i>3.1 Overlaps in disease-associated genes.....</i>	<i>36</i>

3.2	<i>Overlaps in key dysfunctional pathways</i>	38
4	Discussion	44
4.1	<i>Overview</i>	44
4.2	<i>Gene and pathway overlap</i>	44
4.3	<i>Limitations and future work</i>	52
4.4	<i>Conclusion</i>	53
5	References	54
Appendix		68
	<i>Appendix A: scripts for data analysis</i>	68

List of figures

Figure 1- The process of neurodegeneration. The changes from a normal brain to a brain with NDD with reduced brain volume because of neuronal degeneration and subsequent loss of neurons. Reproduced from Pardillo-Díaz et al (2022) 11

Figure 2- The disease courses for MS. The x axis shows time, and the y axis shows disease progression. The red line shows the disease course for CIS, which is one isolated episode, lasting at least 24 hours. The first blue line shows disease progression for RRMS, which is a pattern of relapse followed by complete recovery, the second blue line shows RRMS followed by periods of partial recovery and residual disability. RRMS is the most common disease onset seen in ~85% of cases. The first pink line shows disease course for PPMS, showing a steadily progressively disease, the second pink line shows a progressive disease with varied rates of progression. PPMS onset is seen in ~5-15% of MS cases. The green line shows SPMS, which is typically diagnosed after RRMS, from bouts of relapse and recovery, the disease gets progressively worse in later stages with fewer or no recovery periods. Drawn in Biorender.com 16

Figure 3- GWAS workflow. Cases and controls are differing phenotypes, individuals from these groups have their genomes tested against a set of SNPs. Variants are represented by a data point on the Manhattan plot, the x-axis corresponds to the chromosome and the y-axis to the P-value. More significant associations are represented as taller peaks. A dashed line signifies the decided P-value threshold. Reproduced from Leiserson et al (2013). 17

Figure 4- AD Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the red dotted line are statistically, those above the blue line show a suggestive association. The variants mapped to APOE showed highest significance. Reproduced from Lambert et al (2013) and Kunkle et al (2019). 18

Figure 5- PD Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the red dotted line are statistically significant. This study mapped the most significant variants to SNCA. Reproduced from Rodrigo and Nyholt (2021). 19

Figure 6- ALS Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the dotted line are statistically significant. The most significant variants are mapped to the C9orf72 gene. Reproduced from Van Rheenen et al (2021). 20

Figure 7- LBD Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the dotted line are statistically significant. The most significant variants are mapped to the APOE gene, followed by GBA and SNCA. Reproduced from Chia et al (2021). 21

Figure 8- HD Manhattan plot. The x-axis is chromosome number, and the y-axis is the P-value. Each dot on the plot represents a genetic variant. Those above the red dashed line are considered statistically significant. This study by on the modifiers of HD showed FAN1 was most significantly associated. Reproduced from Lee et al (2022). 22

Figure 9- MS Manhattan plot. The x-axis denotes the chromosome number and the y-axis the P-value. Each point on the plot is a genetic variant, those above the red line have achieved statistical significance. The highest peak and therefore most significant are variants mapped to the MHC. Reproduced from Patsopoulos and de Bakker (2011.) 23

Figure 10- Symptoms across the diseases. Different cognitive, motor, emotional and behavioural symptoms can be observed across the diseases. In AD, LBD and some PD cases cognitive symptoms are seen. It may also be observed across HD, ALS and MS to a milder degree. Motor symptoms are prevalent across the diseases, but in AD only in the later disease stages. In PD and LBD, it is parkinsonian symptoms, in HD and MS it is muscle weakness, spasticity and involuntary movements, in MS it can be in paralysis. The most common emotional symptoms across the diseases are depression and anxiety. Created in Biorender.com 25

Figure 11- Flowchart of methodology. The downstream analysis outlines and visualised. Drawn in Biorender.com 32

Figure 12- Pairwise overlapping genes across AD, PD, ALS, LBD and MS. Significant overlapping genes across the diseases pairwise, visualised with their percentage similarity calculated using online tool Venny (2.1.0). The number of overlapping genes and percentage similarity were: AD-PD (n = 5, 1.7%)(A), AD-ALS (n = 7, 1.7%)(B), AD-LBD (n = 7, 3.4%)(C), AD-MS (n = 7, 1.2%)(D), PD-ALS (n = 9, 2.7%)(E), PD-LBD (n = 13, 10.3%)(F), PD-MS (n = 7, 1.3%)(G), ALS-LBD (n = 2, 0.8%)(H) and ALS-MS (n = 24, 3.9%)(I). No overlaps were identified between HD and the other diseases or between LBD and MS. See Figure 2 for the full list of overlapping genes corresponding to the Venn diagrams in this image. Created in BioRender.com 36

Figure 13- Pairwise overlapping genes between AD, PD, ALS, LBD and MS. Significant genes overlapping between of AD-PD (n = 5)(A), AD-ALS (n = 7)(B), AD-LBD (n = 7)(C), AD-MS (n = 7)(D), PD-ALS (n = 9)(E), PD-LBD (n = 13)(F), PD-MS (n = 7)(G), ALS-LBD (n = 2)(H) and ALS-MS (n = 24)(I), corresponding to the Venn diagrams displayed in Figure 1. Created in BioRender.com 37

Figure 14- Overlap across more than 2 diseases. A- Showing the BTNL2, HLA-DQA1, HLA-DQA2 and HLA-DRA genes overlapping across AD, PD, ALS and MS, with HLA-DRB5 being exclusive to AD, PD and ALS and HLA-DRB1 to AD, PD and MS. B- Showing the DGKQ and TMEM175 genes shared across AD, PD and LBD 37

Figure 15- Semantic similarity scores across the diseases pairwise. BMA semantic similarity scores showing similarity between two sets of GO terms between the diseases pairwise, generated via R package GOSemSim. Scored from 0-1 with 0 being no semantic similarity and 1 denoting complete identity, AD-MS and PD-ALS showed a relatively high semantic similarity of 0.674 and 0.61, respectively. Conversely, ALS-HD, PD-HD and ALS-MS displayed relatively low scores with their respective semantic similarity scores being below 0.3. Created in Biorender.com 38

List of tables

<i>Table 1- Input data for each disease. A column for the disease included in this study, the study from which the summary statistics were obtained from and the number of cases and controls in the original study.</i>	30
<i>Table 2- Pathway overlap between diseases every 2-ways. Each disease compared pairwise, with its corresponding number of overlapping GO terms with a semantic similarity score > 0.8, ranked from the highest number of overlapping pathways to lowest. The greatest number of GO term overlaps can be observed between AD-MS (n = 141). PD-HD and PD-ALS were excluded as there were no overlapping GO terms. GOSemSim was used in R to assess semantic similarity.</i>	39
<i>Table 3- Pathway overlap between diseases every 3 ways. Each disease compared every 3 ways, with its corresponding number of overlapping GO terms with a semantic similarity score > 0.8, ranked from the highest number of overlapping pathways to lowest. The greatest number of overlaps can be seen in the disease trio AD-PD-MS, with 25 shared pathways. GOSemSim was used in R to assess semantic similarity.</i>	40
<i>Table 4- Pathway overlap between diseases every 4-ways. Each disease compared every 4-ways, with its corresponding number of overlapping GO terms with a semantic similarity score > 0.8, ranked from the highest number of overlapping pathways to lowest. The only 4-way overlap was between AD-PD-ALS-MS with 12 shared GO terms between them. Other 4-way comparisons lacked overlapping pathways and were not reported. GOSemSim was used in R to assess semantic similarity</i>	41
<i>Table 5- The top pathways across the diseases based on the number of diseases in which the term appears. The columns show the GO term ID, its corresponding term and a colour pattern of the diseases in which the term is present. recombination was seen across PD, HD and MS. GO terms were enriched by R package org.Hs.eg.db (3.17.0).</i>	43

Abbreviations

- AD = Alzheimer's disease
ALS = Amyotrophic lateral sclerosis
APC = Antigen presenting cell
A β = Amyloid-beta
BED = Browser Extensible Data
BMA= Best Matched Average
BP = Biological Processes
CIS = Clinically Isolated Syndrome
CNS = Central nervous system
DAG = Diacylglycerol
DGK θ = Diacylglycerol kinase theta protein
EOAD = Early-onset AD
GO = Gene ontology
GWAS = Genome-wide association study
HD = Huntington's disease
HLA= Human leukocyte antigen
iNOS = the inducible form of NO synthase enzymes
LB = Lewy bodies
LBD = Lewy body Dementia
LOAD = Late-onset AD
MHC = Major histocompatibility complex
MS = Multiple sclerosis
MSN = Medium-sized spiny neurons
NDD = Neurodegenerative diseases
NFT= Neurofibrillary tangles
NO = Nitrogen species
PA = Phosphatidic acid
PD = Parkinson's disease
PDD = Parkinson disease dementia
PKC = Protein kinase C
PPMS = Primary-progressive MS

ROS = Reactive oxygen species

RRMS = Relapse-remitting MS

SNP = Single nucleotide polymorphisms

SPMS = Secondary-progressive MS

SV = Synaptic vesicle

1 Introduction

1.1 Neurodegenerative diseases

Neurodegenerative diseases (NDD) are a group of disorders affecting neurons in the brain resulting in their progressive atrophy. Neurons are specialised cells of the nervous system which transmit chemical and electrical signals from the brain to the body, coordinating bodily activities. Neuronal communication is critical for nervous system function, including sensory processing, motor control, memory and cognition. Adult neurogenesis occurs in very limited capacity with adult neural stem cells found only in the subventricular zone and hippocampal dentate gyrus regions of the brains (Braun and Jessberger, 2014). Therefore, when neurons are damaged from NDDs, they cannot replace themselves leading to impairments in cognition, memory, mobility, speech and other essential functions (Figure 1).

The NDDs explored in this paper are: Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), Lewy body dementia (LBD), Huntington's disease (HD) and multiple sclerosis (MS).

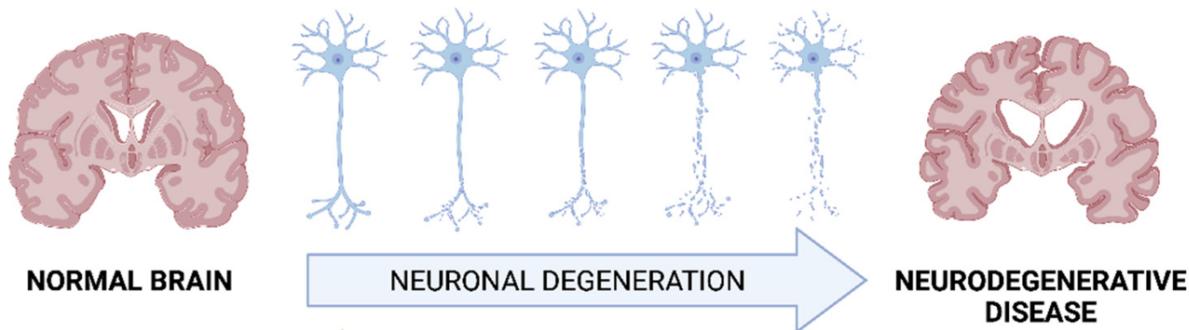


Figure 1- The process of neurodegeneration. The changes from a normal brain to a brain with NDD with reduced brain volume because of neuronal degeneration and subsequent loss of neurons. Reproduced from Pardillo-Díaz et al (2022)

1.1.1 Alzheimer's disease

AD accounts for 60%-70% of all dementias and is the most prevalent type of NDD (WHO, 2023). Currently, it affects 1 in 14 people over the age of 60 which is predicted to triple by

2050 (Nichols *et al.*, 2022). AD results in impairment of episodic memory, followed by aphasia, apraxia and agnosia and behavioural changes (Blennow, Leon and Zetterberg, 2006).

The hallmark pathological features of AD are amyloid plaques (APs) and neurofibrillary tangles (NFTs). APs are deposits of a peptide called amyloid-beta (A β). A β is formed by the cleavage of transmembrane protein, amyloid precursor protein by enzymes β -secretase and γ -secretase. The cleavage forms the A β peptide which tends to misfold and clump with other peptides forming oligomers which can aggregate into plaques in the extracellular space (LaFerla, Green and Oddo, 2007). Accumulation of plaques in the hippocampus, amygdala and cerebral cortex regions of the brain stimulate microglia and astrocytes leading to inflammation (Breijyeh and Karaman, 2020).

NFTs are deposits of insoluble, misfolded, hyperphosphorylated tau protein within the neuronal cytoplasm (Youssef, 2018). In an AD brain, the phosphorylation level of tau was 3-4-fold higher than in a healthy brain (Gong and Iqbal, 2008). Tau provides structure to microtubules, but when hyperphosphorylated, it detaches from the microtubule and polymerises forming NFTs. This can cause obstruction in communication within the neuron and signal processing (DeTure and Dickson, 2019).

1.1.2 Parkinson's disease

PD affects around 10 million people worldwide, with men at a slightly higher risk than women (Marras *et al.*, 2018). Most PD patients are over 60 years old but 10% are under 50. In a report by Parkinson's UK, the prevalence of PD is estimated to increase by 18% between 2018-2025, expected to effect 1 in 30 individuals in the UK (Ohene, 2018).

PD is progressive neurodegenerative condition resulting from the loss of dopaminergic neurons in the brain, specifically in the substantia nigra pars compacta, causing associated motor symptoms such as bradykinesia, muscular rigidity, resting tremor, posture instability and shuffling gait (Sveinbjornsdottir, 2016). Non-motor symptoms can include pain, depression, sleep disturbances, cognitive impairment and genitourinary problems (Chaudhuri, Healy and Schapira, 2006). The neurodegeneration of dopaminergic neurons is correlated with Lewy body (LB) formation which is composed of α -synuclein as neurotoxic soluble

oligomers. This is theorised to cause pathogenesis by disrupting nuclear membrane integrity, releasing a-synuclein aggregates that promotes histone release and inflammation, spreading a-synuclein to neighbouring cells, leading to cell death(Srinivasan *et al.*, 2021).

1.1.3 Amyotrophic lateral sclerosis

ALS is the most common type of motor neuron disease and affects 1.9-6 people per 100,000. ALS is more likely to occur in men than women and a higher incidence of the disease was found in developed countries compared to developing countries (Xu *et al.*, 2020).

ALS is a fatal disease resulting in neurodegeneration of upper and lower motor neurons in the brain and spinal cord. Loss of motor neurons leads to muscle weakness, atrophy and eventual paralysis. ALS has two presentations: spinal onset affecting approximately 70% of ALS patients, and bulbar onset characterised by problems with speech (Shellikeri *et al.*, 2017).

One feature of ALS is protein aggregation within neurons, specifically TDP-43, FUS and SOD1. Protein aggregates may be released into extracellular space and taken up by neighbouring cells hampering their cellular processes (Cicardi *et al.*, 2021). A potential disease mechanism is excitotoxicity which leads to calcium influx into the neurons resulting in increased nitric oxide formation (Wijesekera and Nigel Leigh, 2009). Other mechanisms observed in ALS include, mitochondrial dysfunction, neuroinflammation, oxidative stress, impaired axonal transport and neurofilament aggregation (Wijesekera and Nigel Leigh, 2009).

1.1.4 Lewy body dementia

LBD is the second most common dementia, accounting for 20-30% of cases and is reported to effect males more than females with a ratio of 4:1 (Latimer and Montine, 2021). It is a progressive neurodegenerative disorder that is often misdiagnosed due to its clinical overlap with AD and PD. Symptoms include psychosis specifically visual hallucinations, cognitive impairment, motor parkinsonism and sleep behaviour disorders (Chin, Teodorczuk and Watson, 2019).

LBD is distinguished by the presence of LB and Lewy-neurites in the brainstem, limbic system and cortical regions. LB and Lewy-neurites are the result of abnormal a-synuclein, with abnormal solubility, that prompts the production of oligomeric species and aggregates into fibrils. This results in membrane disruption, histone release, inflammation and spread of a-synuclein aggregates, thought to lead to progressive neurodegeneration (Garcia-Esparcia *et al.*, 2017). LBD is encompassed by two clinical entities, Parkinson disease dementia (PDD) and dementia with Lewy bodies (DLB). The distinction between the two is in the primary order of symptom onset. In PDD, dementia begins at least a year after PD onset whereas in DLB, dementia precedes or co-occurs with parkinsonism (Smirnov *et al.*, 2020).

1.1.5 Huntington's disease

HD is a rare monogenic neurodegenerative disorder, with an autosomal-dominant inheritance pattern. Global prevalence is around 2.7 per 100,000. Symptoms include involuntary jerky movements, altered personality, cognitive impairment and psychiatric symptoms (Chial, 2008).

HD is caused by an expansion of CAG repeats in the coding region of the HTT gene, leading to the production of mutated huntingtin protein that tends to accumulate in neuronal cytoplasm and nuclei (Jimenez-Sanchez *et al.*, 2017). The pathogenic HTT mechanisms include dysregulation of transcription factors, impaired mitochondrial pathways, altered protein homeostasis, increased presence of aggregates that interfere with other cellular factors, affected axonal trafficking of vesicles, organelles and neurotransmitters, synaptic plasticity failure and glial activation (Gatto *et al.*, 2020).

The striatum and cortex regions of the brain are primarily affected in HD. Medium-sized spiny neurons (MSNs) which make up around 90% of the striatum are particularly vulnerable to the effects of mutated huntingtin (Albin *et al.*, 1990). While striatal neuronal death may be the underlying cause of late-stage HD symptoms, early deficits are likely associated with cellular and synaptic dysfunction in the cortex (Raymond *et al.*, 2011). The severity of neurodegeneration in different regions of the cortex correlate to manifested symptoms, for example the extent of neuron loss in the primary motor cortex correlates to motor

dysfunction, whereas extent of loss in the cingulate cortex relates to behavioural changes (Thu *et al.*, 2010).

1.1.6 Multiple sclerosis

MS is a progressive, autoimmune and neurodegenerative disease of the central nervous system (CNS). Globally, MS effects ~35.9 people per 100,000, an estimated 30% increase from 2013 to 2020 (Walton *et al.*, 2020). Females are twice as likely to be affected than males, with a 4:1 ratio in some countries (Walton *et al.*, 2020). Symptoms vary between individuals and include spasticity, muscle weakness, bowel and sexual dysfunction, cognitive impairment, pain and fatigue (Crayton and Rossman, 2006).

The current understanding of pathology is that CD4+ T-cells, specifically Th17, infiltrate the CNS through a disrupted blood-brain barrier and cause activation of T-cells, effector cytokines, chemokines and other immune cells (Korn, 2008). This inflammatory response causes damage to myelin and oligodendrocytes, a type of glial cell that produces myelin around the axons of neurons (Patel and Balabanov, 2012). Another observation is the release of toxic reactive oxygen species and nitrogen species which contribute to myelin and axonal damage (Faissner *et al.*, 2019). Lesions form around areas where the myelin sheath has become destroyed. In the most common disease course, attacks are followed by a recovery period during which there is remyelination. However, recovery tends to be partial, usually leaving some residual disability. After repeated episodes, damage becomes irreversible leading to neurodegeneration. There are four disease courses: Clinically Isolated Syndrome (CIS), Relapse-Remitting MS (RRMS), Primary-Progressive MS (PPMS) and Secondary-Progressive MS (SPMS; Figure 2).

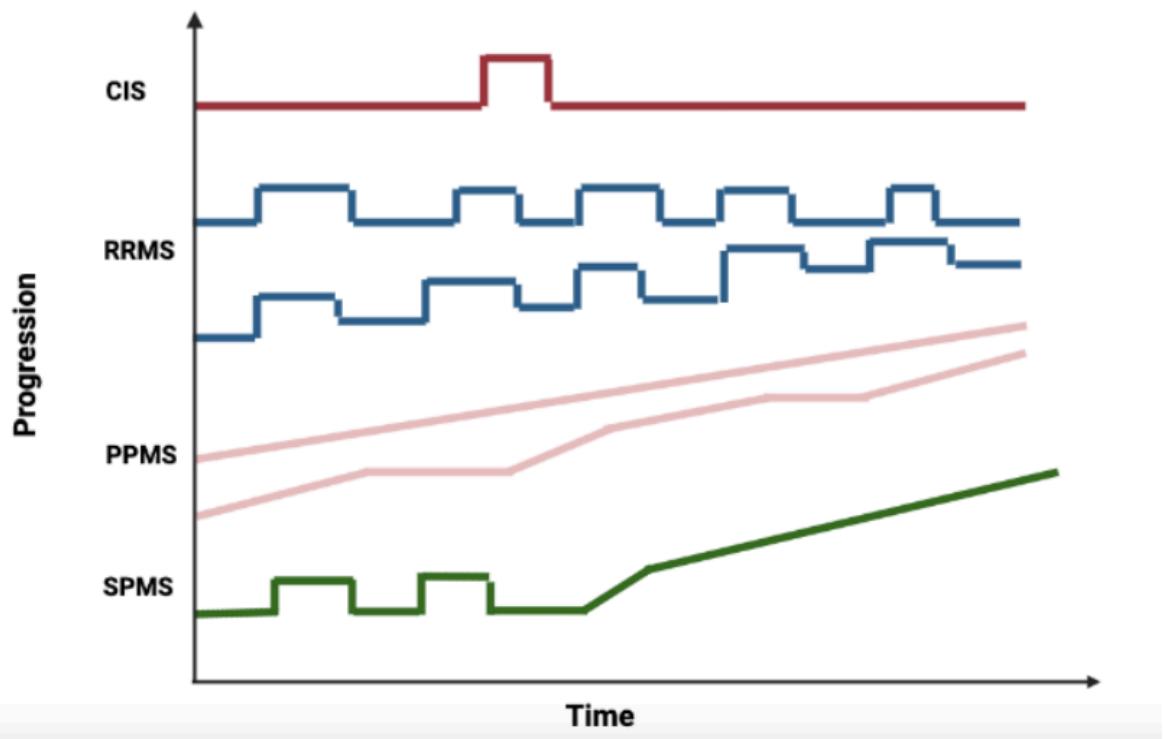


Figure 2- The disease courses for MS. The x axis shows time, and the y axis shows disease progression. The red line shows the disease course for CIS, which is one isolated episode, lasting at least 24 hours. The first blue line shows disease progression for RRMS, which is a pattern of relapse followed by complete recovery, the second blue line shows RRMS followed by periods of partial recovery and residual disability. RRMS is the most common disease onset seen in ~85% of cases. The first pink line shows disease course for PPMS, showing a steadily progressively disease, the second pink line shows a progressive disease with varied rates of progression. PPMS onset is seen in ~5-15% of MS cases. The green line shows SPMS, which is typically diagnosed after RRMS, from bouts of relapse and recovery, the disease gets progressively worse in later stages with fewer or no recovery periods. Drawn in Biorender.com

1.2 Genetics

Advances in identification of disease-causing genes and risk loci have led to invaluable understanding of NDD mechanisms and pathways (Uffelmann *et al.*, 2021).

Genome-wide association studies (GWAS) are an approach used to map the genetic framework of diseases/traits by identifying statistically significant genetic variants associated with a phenotype of interest. A set of variants, usually single nucleotide polymorphisms (SNPs) are tested against the genomes of many individuals belonging to an ancestrally similar population but with differing known phenotypes (Figure 3).

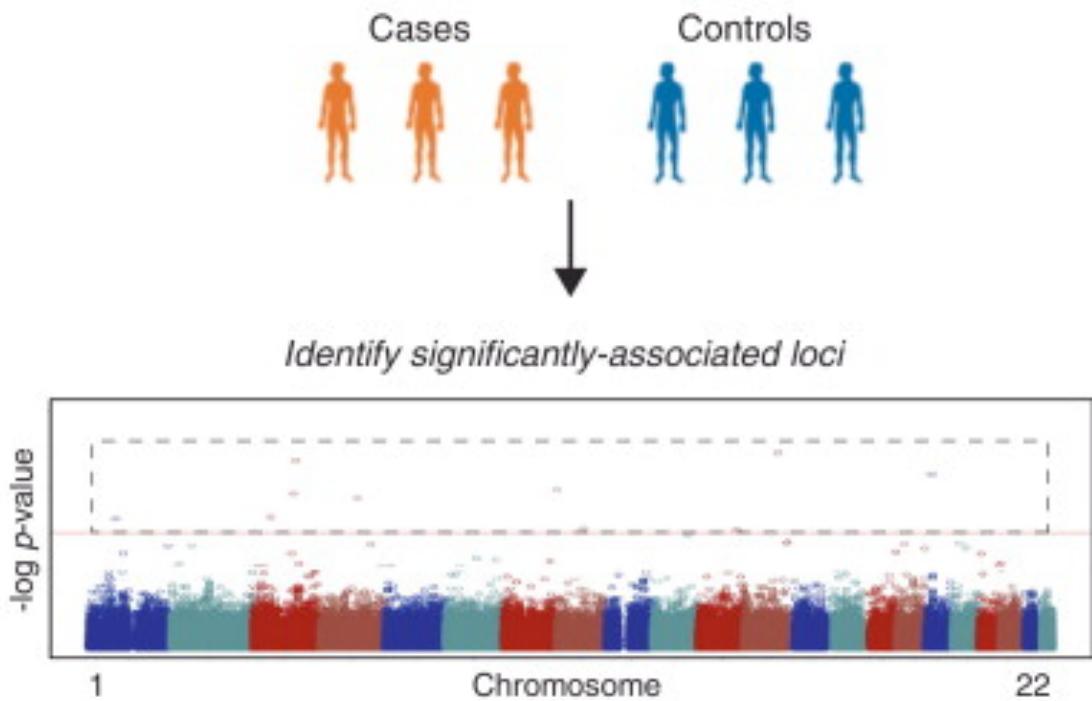


Figure 3- GWAS workflow. Cases and controls are differing phenotypes, individuals from these groups have their genomes tested against a set of SNPs. Variants are represented by a data point on the Manhattan plot, the x-axis corresponds to the chromosome and the y-axis to the P-value. More significant associations are represented as taller peaks. A dashed line signifies the decided P-value threshold. Reproduced from Leiserson et al (2013).

1.2.1 Genetics of Alzheimer's disease

Early-onset AD (EOAD) occurs before the age of 65 and accounts for around 10% of AD cases. Highly penetrant mutations in the following genes cause EOAD: APP, PSEN1 and PSEN2 (Cacace, Sleegers and Van Broeckhoven, 2016). APP encodes amyloid precursor protein, a membrane glycoprotein. Mutations lead to an increase in the amyloidogenic pathway processing by γ and β secretase leading to A β production (Haass *et al.*, 2012). PSEN1 and PSEN2 disease gene mutations also lead to an increase in A β production of 1.2-3-fold (Haass and Strooper, 1999).

Late-onset AD (LOAD) is genetically complex and results from a combination of genetic and environmental factors. The main genetic risk factor for LOAD is the $\epsilon 4$ allele of the APOE gene (Figure 4). One inherited copy of the $\epsilon 4$ allele increased risk by approximately 4-fold and two copies by over 10-fold (Tanzi, 2012). In contrast the $\epsilon 2$ allele is thought to exert

protective effects (Corder *et al.*, 1994). Differential effects of APOE alleles on disease risk is likely mediated by differential effects on A β accumulation (Vergheze, Castellano and Holtzman, 2011).

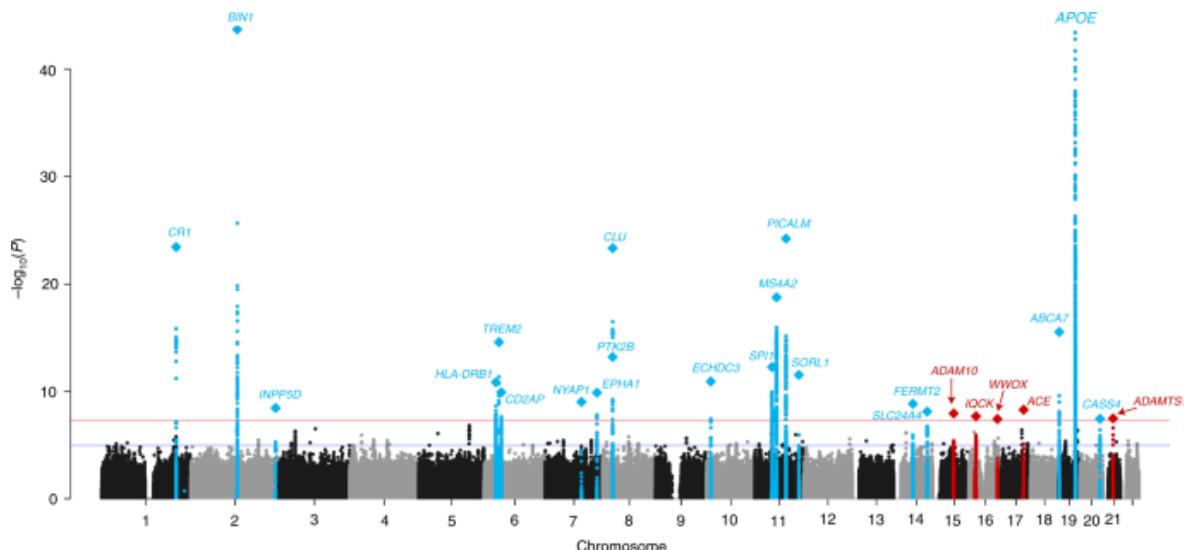


Figure 4- AD Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the red dotted line are statistically, those above the blue line show a suggestive association. The variants mapped to APOE showed highest significance. Reproduced from Lambert *et al* (2013) and Kunkle *et al* (2019).

1.2.2 Genetics of Parkinson’s disease

There has been identification of several monogenic forms of PD and numerous genetic loci associated with increased disease risk. Approximately 5-10% of PD cases are monogenic, a result of a highly penetrant rare mutation inherited in an autosomal-dominant or recessive pattern. However, most PD cases are caused by a combination of factors including, common genetic variants with small to moderate effect size and their possible interaction, environmental factors, lifestyle exposure and epigenetic factors (Lill, 2016).

Mutations in SNCA, LRRK2 and VPS35 genes are known to cause monogenic forms of PD and are inherited in an autosomal-dominant pattern. Parkin, PINK1 and DJ-1 genes also cause monogenic forms of PD but are inherited in an autosomal-recessive pattern (Figure 5). Other mutations in LRRK2 and GBA genes are strong risk factors but do not have complete penetrance (Bonifati, 2014). Common variants with smaller effect size have been identified by GWAS with 90 risk associations found across 78 genomic loci (Nalls *et al.*, 2019).

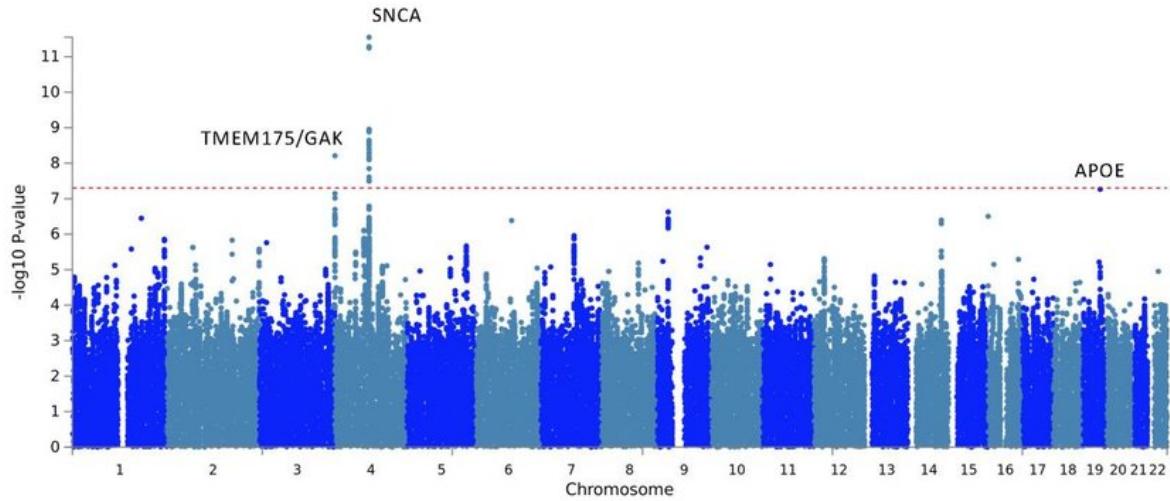


Figure 5- PD Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the red dotted line are statistically significant. This study mapped the most significant variants to SNCA. Reproduced from Rodrigo and Nyholt (2021).

1.2.3 Genetics of amyotrophic lateral sclerosis

Over 50 genes are associated with familial ALS, which comprises 10% of ALS cases (Boylan, 2015; Figure 6).

Over 185 SOD1 variants have been identified in ALS (Yamashita and Ando, 2015). SOD1 encodes a superoxide dismutase enzyme which binds copper and zinc to form dimers that protect cells from oxidative stress (McCord and Fridovich, 1969). Recent studies revealed a toxic gain-of-function increasing excitotoxicity, oxidative stress and endoplasmic reticulum stress (Hayashi, Homma and Ichijo, 2016).

Over 50 FUS mutations have been linked to ALS. FUS is involved in regulating gene expression and RNA processing, most mutations affect nuclear localisation signals of the protein causing formation of FUS-positive inclusions (An *et al.*, 2019). Strong evidence points to a toxic gain-of-function mediated by cytoplasmic FUS aggregates (Mejzini *et al.*, 2019). Similarly, disease mechanisms in TARDBP-associated ALS parallels that of FUS. TARDBP encodes the TDP-43 protein, involved in gene expression and RNA processing. In ALS, TDP-43 mislocalisation occurs from the nucleus to the cytoplasm causing protein

aggregation. Depletion from the nucleus results in compensatory upregulation of TDP-43 synthesis. Both loss and overexpression are implicated in disease (Mejzini *et al.*, 2019).

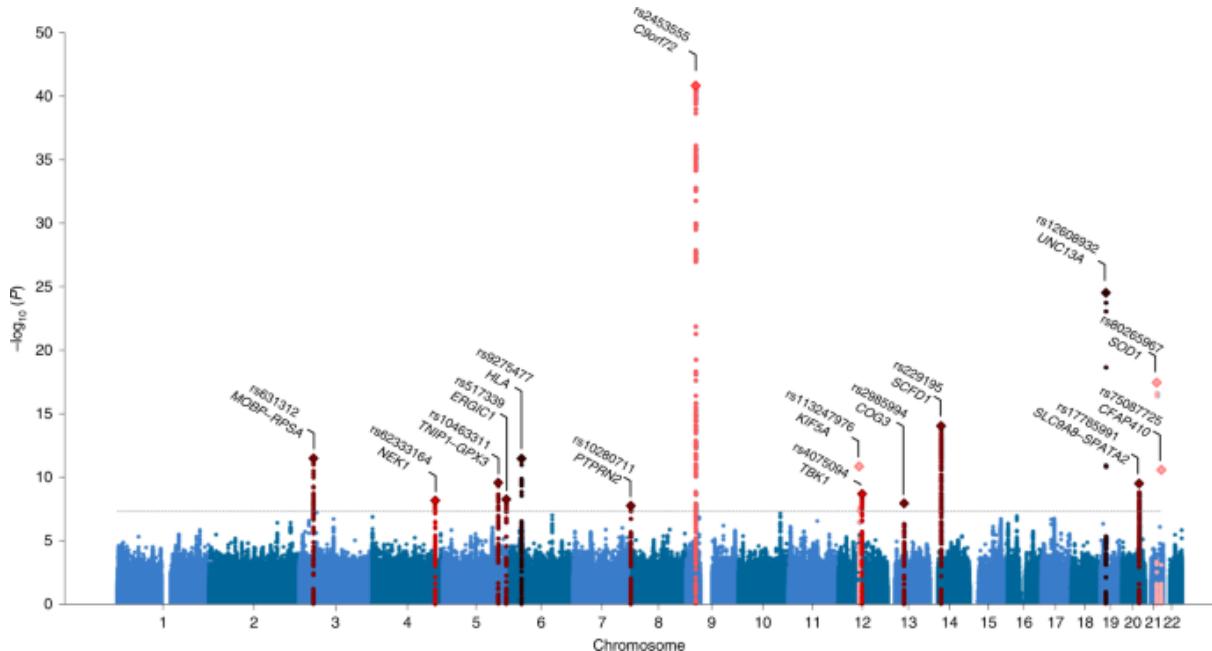


Figure 6- ALS Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the dotted line are statistically significant. The most significant variants are mapped to the C9orf72 gene. Reproduced from Van Rheenen *et al* (2021).

1.2.4 Genetics of Lewy body dementia

The SNCA gene was first implicated in LBD in families showing mixed phenotypes of dementia and parkinsonism (Orme, Guerreiro and Bras, 2018). A study by Guella *et al* (2016) found risk of dementia was associated with mutations at the 5' locus of the SNCA gene while risk of parkinsonism was associated with mutations at the 3' locus.

The APOE e4 allele has been associated with an increased risk of LBD (Figure 7). It is theorised that the allele contributes to β-amyloid deposition and LB pathology through an amyloid cascade hypothesis like that seen in AD (Hardy and Higgins, 1992). Conversely, the APOE e4 allele was associated with a higher LB count in LBD cases independent of AD pathology, suggesting the allele increases the risk of LBD by directly affecting build-up of LB in neurons (Dickson *et al.*, 2018).

GBA mutations are strongly associated with LBD (Figure 7). GBA encodes glucocerebrosidase (GCase), an enzyme responsible for cleaving glucosylceramide and glucosylsphingosine lipids (Smith and Schapira, 2022). Evidence from cell models theorise that mutations reduce GCase levels and activity due to disrupted trafficking of GCase to the Golgi leading to accumulation of glucosylceramides and impaired α -synuclein degradation (Mazzulli *et al.*, 2011).

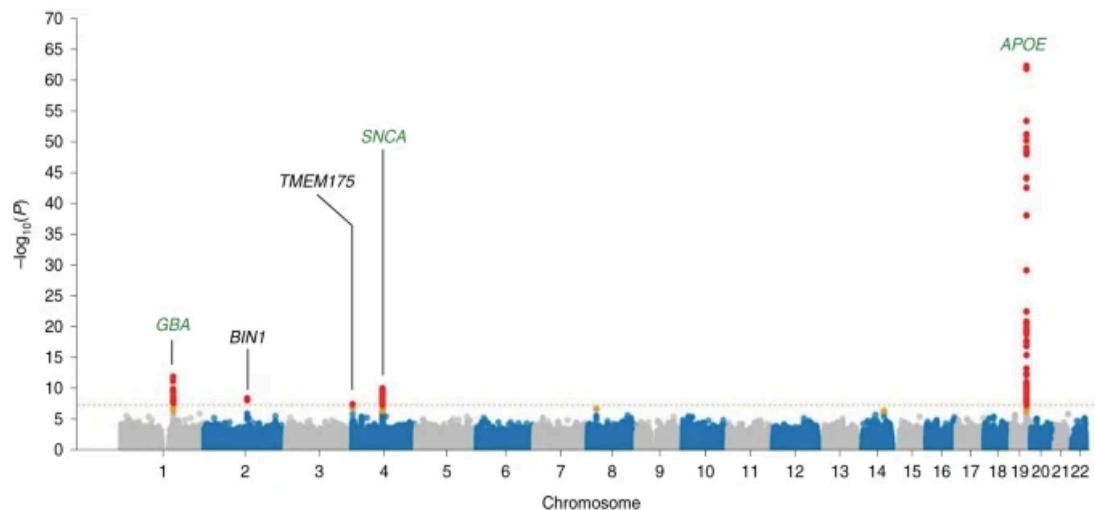


Figure 7- LBD Manhattan plot. The y-axis denotes chromosome number and the x-axis the P-value. Each point on the plot is a genetic variant, those above the dotted line are statistically significant. The most significant variants are mapped to the APOE gene, followed by GBA and SNCA. Reproduced from Chia et al (2021).

1.2.5 Genetics of Huntington's disease

The sole disease-causing mutation in HD is a CAG repeat expansion in the coding portion of the HTT gene (Figure 8). Inheritance is primarily autosomal-dominant, however around 3-4% of cases occur sporadically (Batino *et al.*, 2021). The HTT gene encodes the huntingtin protein. The number of CAG repeats can vary between affected individuals, with a higher number of repeats generally correlating to earlier disease onset and severity. Mutant huntingtin with longer polyglutamine tracts are more prone to misfolding, aggregation and toxicity.

A CAG repeat length ≥ 40 represents a range that will develop into HD, with complete penetrance and clinical expression (Gatto *et al.*, 2020). No confirmed HD cases have been reported in individuals with ≤ 26 CAG repeats (Nance, 2017). Interestingly, the sex of the

transmitting parent is a major determinant of intergenerational changes. When the mutation is inherited paternally, the CAG repeat length seems to significantly increase between generations compared to maternal inheritance which demonstrates intergenerational stability, this phenomenon is known as paternal anticipation (Gupta and Kushwaha, 2019).

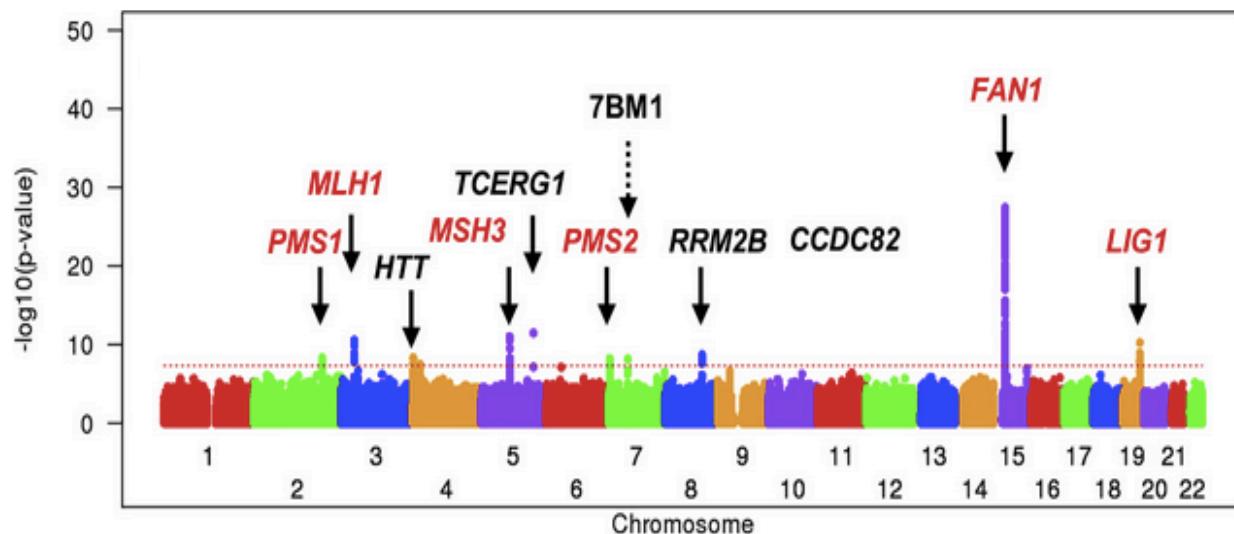


Figure 8- HD Manhattan plot. The x-axis is chromosome number, and the y-axis is the P-value. Each dot on the plot represents a genetic variant. Those above the red dashed line are considered statistically significant. This study by Lee et al (2022) showed *FAN1* was most significantly associated.

1.2.6 Genetics of multiple sclerosis

MS is thought to be influenced by a combination of genetic and environmental factors. While initial studies found associations to class I *HLA-A* and *HLA-B* alleles, studies with increased power found the strongest susceptibility mapped to the HLA-DRB1 gene in the class II MHC region (Didonna and Oksenberg, 2017).

The largest genetic study of MS identified 233 genome-wide loci associated with disease susceptibility, with 200 loci located outside the MHC (International Multiple Sclerosis Genetics Consortium, 2019; Figure 9). Most MS variants fall within intronic or intragenic regions making it difficult to map them to mechanisms and pathways (Patsopoulos, 2018). The risk of MS is ~0.2% in the general population. Having affected siblings increases risk to 2-4%, and for monozygotic twins, it rises to 30%. Non-related family members have a similar

risk to the general population. This confirms genetic significance in MS development, but relative risk does not reach 100% even in monozygotic twins suggesting other factors may influence susceptibility (Didonna and Oksenberg, 2017).

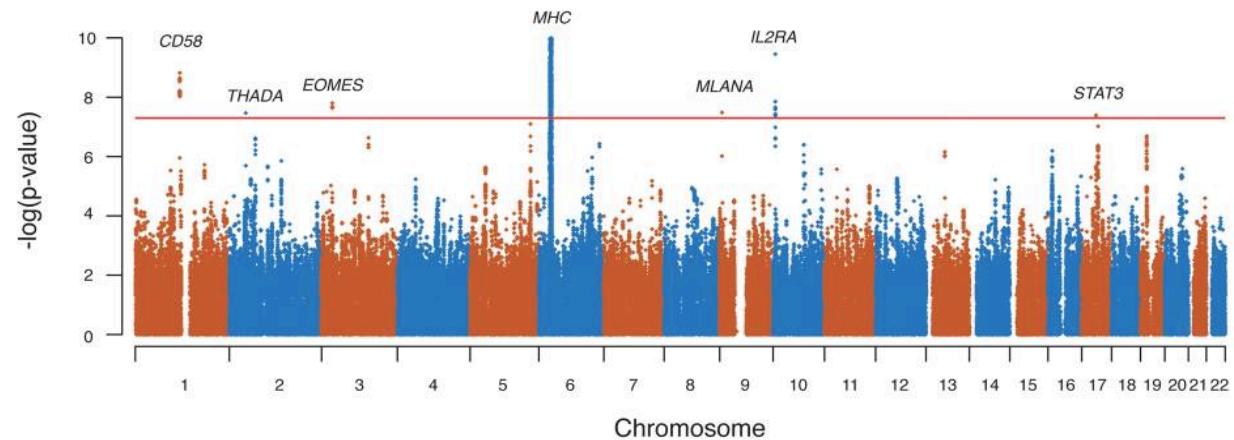


Figure 9- MS Manhattan plot. The x-axis denotes the chromosome number and the y-axis the P-value. Each point on the plot is a genetic variant, those above the red line have achieved statistical significance. The highest peak and therefore most significant are variants mapped to the MHC. Reproduced from Patsopoulos and de Bakker (2011.)

1.3 Commonalities across neurodegenerative diseases

While these diseases are considered separate diseases, there are significant overlaps in risk factors, genetic influences, symptoms exhibited, pathology and disease mechanisms. This overlap and distinction will be discussed in the following section.

1.3.1 Shared risk factors

Age is a significant risk factor for NDD's. The brain's post-mitotic cells are vulnerable to the effects of aging and DNA damage, promoting cellular senescence and inflammation, which exacerbate neurodegeneration (Madabhushi, Pan and Tsai., 2014; Hou et al., 2019). In addition, aging neurons often exhibit mitochondrial damage and dysfunction, in both healthy aging and in NDDs (Hou et al., 2019).

There is evidence of pollution, pesticides, heavy metals and head trauma increasing the risk of AD, PD, ALS, and LBD. Exposure to pollution can lead to chronic oxidative stress (Moulton and Yang, 2012), while heavy metals and pesticides can dysregulate lysosomal and mitochondrial function, enhancing spread of misfolded proteins and potentially triggering

inflammation (Wang *et al.*, 2021). Additionally, cigarette smoking may contribute to the risk of AD, ALS and MS by increasing oxidative stress or through direct neurotoxic effects on motor neurons in ALS (Wang *et al.*, 2011). Interestingly, some studies suggest a protective effect of cigarette smoking in PD with one study observing a 20% decreased risk in former smokers and a halved risk in current smokers (Gallo *et al.*, 2019). A dysfunctional blood-brain barrier is seen in many diseases including AD, PD and MS, leading to increased permeability that can allow bacteria and viruses to enter the brain (Sait *et al.*, 2021).

1.3.2 Shared symptoms

While the diseases are clinically distinct, they exhibit similar symptoms. AD primarily presents cognitive decline; LBD also shows cognitive decline but to a less severe degree than AD (Smirnov *et al.*, 2020). Dementia in PD is less severe than in AD and LBD and affects ~ 40% of cases (Fang *et al.*, 2020). ALS, HD, and MS may also display mild cognitive decline. Impaired movement is common in PD, LBD, ALS, HD, and MS. PD and LBD manifest as parkinsonian symptoms. HD includes involuntary movement, MS has muscle weakness, and ALS displays severe muscle atrophy and paralysis. In late-stage AD, motor dysfunction can also be observed (Wirths and Bayer, 2008). Emotional disruptions occur in NDDs. AD patients often display apathy, depression, anxiety, agitation, aggression, hallucinations, or delusions (Lyketsos *et al.*, 2002). Similar symptoms exist across PD, ALS, LBD, HD, and MS with depression and anxiety being the most common. Symptoms across the diseases can be seen in Figure 10.

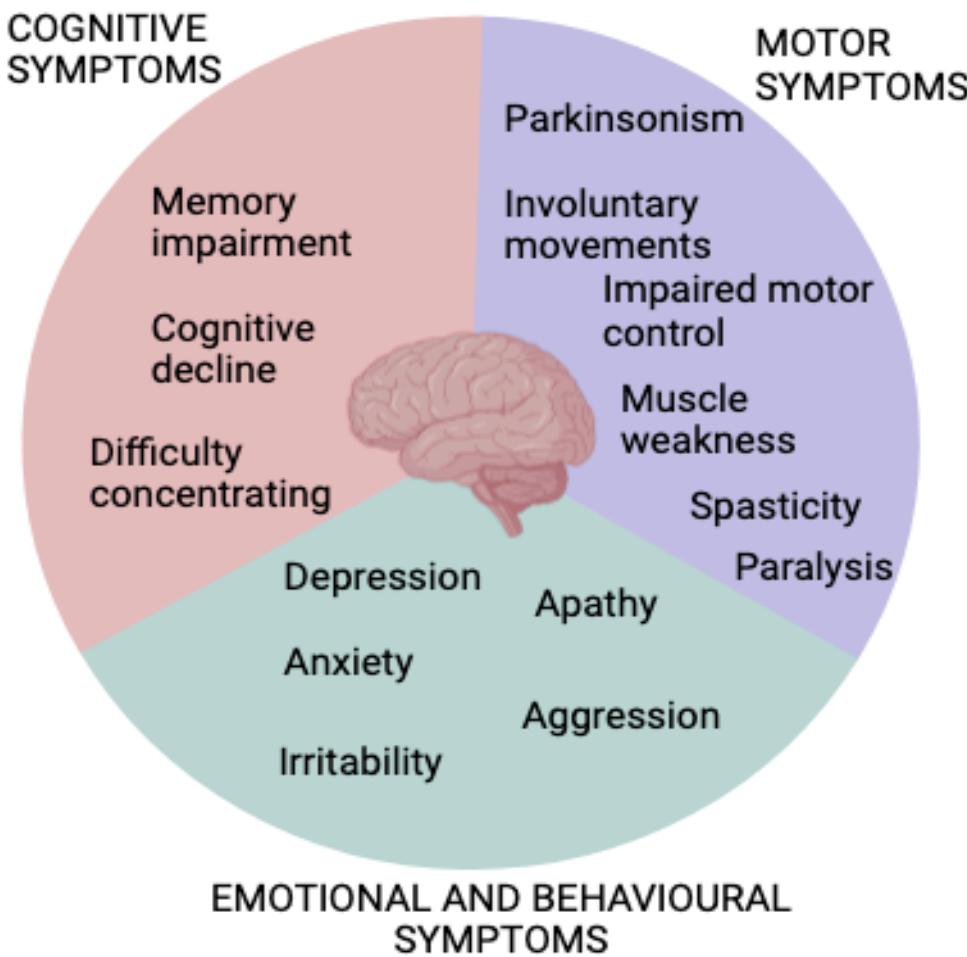


Figure 10- Symptoms across the diseases. Different cognitive, motor, emotional and behavioural symptoms can be observed across the diseases. In AD, LBD and some PD cases cognitive symptoms are seen. It may also be observed across HD, ALS and MS to a milder degree. Motor symptoms are prevalent across the diseases, but in AD only in the later disease stages. In PD and LBD, it is parkinsonian symptoms, in HD and MS it is muscle weakness, spasticity and involuntary movements, in MS it can be in paralysis. The most common emotional symptoms across the diseases are depression and anxiety. Created in Biorender.com

1.3.3 Shared pathology

Neurodegenerative diseases have common cellular and molecular mechanisms including protein aggregation and inclusion body formation. The formation of misfolded protein aggregates can contribute to cellular proteostatic collapse and dysfunction (Sweeney *et al.*, 2017). The implicated protein differs between diseases; A β and tau in AD, α -synuclein in PD, TDP-43, FUS and SOD1 in ALS, α -synuclein in LBD and mutant HTT in HD, all of which conform into fibrillar structures and toxic oligomeric species with neurotoxic effects. MS is the exception with no known protein aggregation.

While LBD is defined by the presence of α -synuclein aggregates like those seen in PD, many cases of LBD show concurrent AD pathology in the form of APs and tau NFTs (Schumacher *et al.*, 2021). Although α -synuclein is the key underpinning of PD, there can be comorbid associations with A β and tau deposition, with A β being linked to cognitive decline (Lim *et al.*, 2019). However NFTs and astroglia anomalies observed were fewer and less severe compared to AD (Foguem and Manckoundia, 2018). The presence of A β deposits at the cortical level, hippocampus and spinal cord motor neurons have also been described in ALS patients, with amyloid cascade activation in the hippocampus being correlated with cytoplasmic phosphorylated TDP-43 expression (Colletti *et al.*, 2021) (Gómez-Pinedo *et al.*, 2016).

Polyglutamine inclusions in HD have been identified as immunopositive for α -synuclein (Chánez-Cárdenas and Vázquez-Contreras, 2012). Upon discovery, it was speculated α -synuclein may be employed as an addition mediator of polyglutamine toxicity. Accumulation of α -synuclein in polyglutamine inclusions observed in HD post-mortem brains confirmed this hypothesis (Chánez-Cárdenas and Vázquez-Contreras, 2012). In HD mice models, α -synuclein overexpression enhanced onset of tremors while deletion delayed tremors (Corrochano *et al.*, 2012).

1.3.4 Shared disease mechanisms

NDDs also share common causes for protein aggregation, which leads to disruptions in regular cellular function due to the accumulation of these abnormal proteins. Autophagy plays a key role in degradation of abnormal proteins seen in AD, PD, ALS, LBD and HD. The ubiquitin proteasome system and autophagy can both clear ubiquitinated substrates, but autophagy is the only way to clear large aggregates that are too narrow to fit the proteasome chamber (Guo *et al.*, 2017). Many studies have shown dysregulated autophagy is closely linked to NDD's. For example, the amount of autophagic vacuoles is considerably higher in the brain of NDD patients, suggesting impaired maturation of autophagosomes to autolysosomes (Yu *et al.*, 2005). In AD, the maturation of autophagolysosomes and their passage back to the neuronal body is hindered, resulting in accretion of autophagic vacuoles within neurons (Nixon, 2007). In addition, A β peptides can accumulate in autophagosomes

within dystrophic neurites, becoming a reservoir of toxic peptides that has neurotoxic effects (Nixon *et al.*, 2005; Zhang *et al.*, 2002). Lysosomes are structures that contain enzymes and break down the contents of autophagosomes. In PD brain samples, dysfunctional lysosomes and autophagosome accumulation was observed, with lysosome inhibition positively correlating with α -synuclein levels (Dehay *et al.*, 2010). In ALS, some studies suggest that impaired cargo digestion in the lysosome leads to autophagy dysfunction (Sasaki, 2011), while other studies indicate a close relationship between increased autophagosomes and decreased mTOR phosphorylation, a pathway that regulates autophagy (Morimoto *et al.*, 2007). Across AD, PD, ALS, LBD, and HD; regardless of the specific mechanism impairing autophagy, the ultimate consequence remains consistent, a failure to degrade neurotoxic protein aggregates.

Several mechanisms that drive neurodegeneration may be triggered by inflammatory cells and their mediators at various stages of the neurodegenerative cascade. While neuroinflammation is seen across all the diseases, it is a prominent feature of MS. Inflammation in MS is characterised by immune responses involving T-cells, B-cells and myeloid cells. Depending on activation state and specific conditions in the body, these cells can amplify or suppress immune responses. It is suggested that there is an imbalance of pro-inflammatory immune cells and regulatory immune cells in the periphery. This phenomenon is linked to the ability of immune cells performing a phenotype-switch, resulting in a defective function of regulatory cells and thereby increasing infiltration of autoreactive adaptive immune cells into the CNS (Haase and Linker, 2021). Inflammatory mediators are known to enhance or affect several neurodegenerative mechanisms through key downstream mediators, specifically reactive oxygen species (ROS), nitrogen species (NO), and cytokine-mediated neurodegeneration. Activated immune cells can produce ROS contributing to mitochondrial dysfunction and eventual apoptosis (Chitnis and Weiner, 2017). NO made up of the inducible form of NO synthase enzymes (iNOS), is produced in response to inflammatory stimuli. NO can inhibit mitochondrial respiration, enhancing the release of neurotransmitter glutamate, initiating an excessive and uncontrolled influx of calcium ions into neurons causing excitotoxic neuronal death (Tewari *et al.*, 2021). Inflammatory cytokines TNF- α and IL-1 β mediate neuroprotection and plasticity, but when it the presence of iNOS become neurotoxic (Bal-Price, Moneer and Brown, 2002). For example, TNF- α induces oligodendrocyte apoptosis and demyelination, pathology typically seen in MS (Taupin *et al.*, 1997). While these mechanisms are initially employed by the immune system

as a protective measure against protein aggregation, their excessive activation leads to neurotoxicity and a cascade of inflammation, ultimately resulting in neuronal death.

There is evidence of oxidative stress and resulting cellular damage across the NDD's. It is induced by an imbalance involving either excessive ROS production, primarily a product of mitochondrial respiration, or dysfunction of the antioxidant system that neutralise these free radicals. The brain is especially vulnerable to the effects of ROS due to its high oxygen demand, leading to increased cellular respiration and consequent ROS production, and due to its abundance of peroxidation-susceptible lipid cells (Kim *et al.*, 2015). Accruing oxidative stress can cause cellular damage, impaired DNA repair, protein misfolding, glial cell activation, mitochondrial dysfunction and subsequent apoptosis (Federico *et al.*, 2012) (Chen, Guo and Kong, 2012). As oxidative stress damages mitochondrial components, thus impairing mitochondrial function, it can further exacerbate ROS production, creating a positively regulated cycle. The exact mechanisms underlying the disruption of redox balance remains elusive, with potentially varying causative mechanisms across the NDDs.

1.4 Justification for study, hypothesis and aims

Between 2015 and 2050, the proportion of the world's population over 60 years will nearly double from 12% to 22% (WHO, 2022). By 2030, 1 in 6 people in the world will be aged 60 years or over (WHO, 2022). While it is excellent that medical advances have increased the average life expectancy, the caveat is that it comes with an increased number of individuals suffering with NDDs, as age is a primary risk factor. Dementia is the 7th leading cause of global mortality with 1 in 6 individuals today suffering from it, however this is predicted to triple by 2050 (Nichols *et al.*, 2022; WHO, 2023). These diseases have debilitating impact on not only the patient, but can also be emotionally, physically, and financially taxing on their caregivers. This highlights the importance of continued research and exploration of therapeutic options, but this can only be achieved by understanding the genes and pathways that need to be targeted in treatment. Given the numerous overlaps in these diseases that have been explored in the above sections, it can be hypothesised that NDDs, are essentially different manifestations of the same disease, and there will be significant genetic, and pathway overlap across them.

This study posits that understanding overlapping genes and pathways across AD, PD, ALS, LBD, HD and MS, will highlight the areas that can be common therapeutic drug targets. Advances in bioinformatics and its tools have allowed for identification of key genes and pathways observed across these diseases from existing GWAS data, and the visualisation of overlaps to draw conclusions from. The aims of this study are as follows:

- To identify key genes implicated in AD, PD, ALS, LBD, HD and MS using SNPs from GWAS summary statistics.
- To determine disease pathways in AD, PD, ALS, LBD, HD and MS using enrichment analysis.
- To discover which significant mutations, genes and pathways overlap between diseases that could be potential targets for treatment.

This will be achieved by selecting the most comprehensive GWAS for each disease and filtering them to get significant SNPs, these will then be mapped to genes, and those genes to pathways via bioinformatic tools available online and in RStudio. Results will then be manipulated to produce appropriate visualisation from which conclusions regarding overlap across diseases can be drawn.

2 Methods

2.1 Input data

To gather data for the analysis, a literature search was conducted to identify the most recent and comprehensive GWAS for each disease. Data was obtained either from publicly available sources, such as the GWAS catalog (Sollis *et al.*, 2023) or directly from the researchers upon request. The studies for each disease can be seen in Table 1:

Table 1- Input data for each disease. A column for the disease included in this study, the study from which the summary statistics were obtained from and the number of cases and controls in the original study.

Disease	Study	Cases	Controls
AD	Wightman <i>et al</i> (2021)	90,338 (46,613 proxy)	1,036,225 (318,246 proxy)
PD	Nalls <i>et al</i> (2019)	37,688 (18,618 proxy)	1,400,000
ALS	Van Rheenen <i>et al</i> (2021)	29,612	122,656
LBD	Chia <i>et al</i> (2021)	2,591	4,027
HD	Moss <i>et al</i> (2017)	1991	0
MS	IMSGC (International Multiple Sclerosis Genetics Consortium) (2019)	47,429	68,374

The number of SNPs analysed for each disease were: 12,688,339 (AD), 17,510,617 (PD), 10,461,755 (ALS), 7,644,957 (DLB), 7,518,232 (HD) and 8,278,136 (MS).

2.2 Analysis

Data analysis was performed primarily in RStudio (version 2023.06.0+421) and to a small degree in Python Jupyter Lab (3.4.4) via Anaconda Navigator (2.4.0). Tools used in RStudio included tidyverse (2.0.0) (Wickham and Grolemund, 2016) for data wrangling, enrichGO

from the package clusterProfiler (4.8.2) (Wu *et al.*, 2021) for pathway enrichment, org.Hs.eg.db (3.17.0) (Carlson, 2023) for gene ontology (GO) term annotations and GOSemSim (2.26.1) (Yu, 2020; Wang *et al.*, 2007) to calculate semantic similarity between sets of GO terms. In Python, the pandas (1.5.3) (McKinney, 2010) package was used to manipulate datasets into Browser Extensible Data (BED) files for GREAT (Genomic Region Enrichment of Annotations Tool) (4.0.4) (McLean *et al.*, 2010; Tanigawa, Dyer and Bejerano, 2022) analysis, an online tool used for gene mapping. The methodology is outlined in Figure 11.

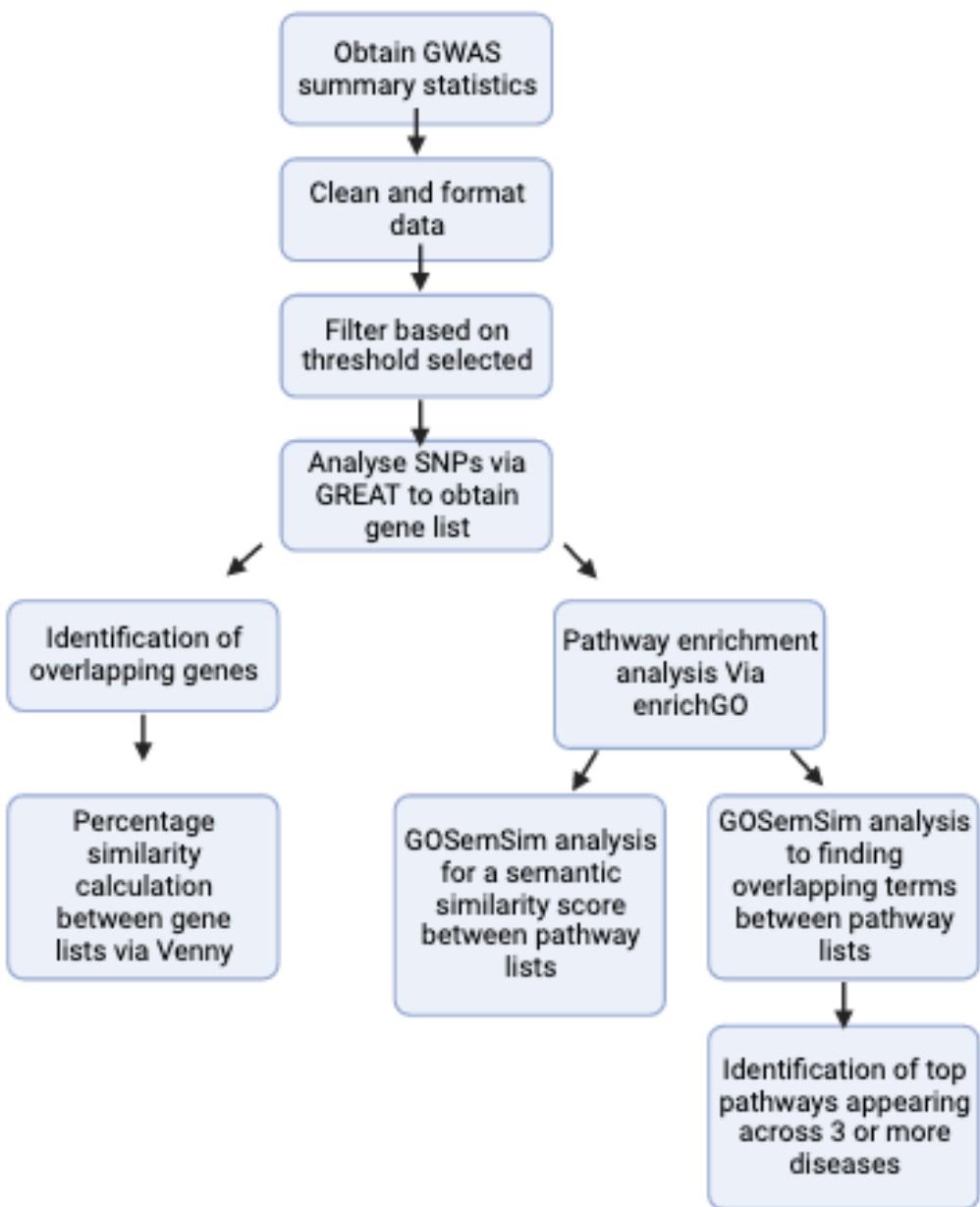


Figure 11- Flowchart of methodology. The downstream analysis outlines and visualised. Drawn in Biorender.com

2.3 Data cleaning and filtering

Before analysis the acquired data was cleaned and standardised in RStudio. Rows with missing values in the P-value or Position column were removed. Columns that were not consistently in all datasets were eliminated, and the remaining columns were reordered and renamed to ensure uniformity across datasets. Retained columns were, ‘Position’ for chromosome number and location, ‘A1’ for the test allele, ‘A2’ for the alternate allele and ‘P’

for the p-value. Datasets were then filtered based on the universal GWAS p-value threshold of 5×10^{-8} , retaining only statistically significant disease-associated SNPs for subsequent analysis. This stringent threshold has been employed over the last decade to establish an association between a common genetic variant and a trait, while accounting for multiple testing (Chen *et al.*, 2021). However, initial analysis of ALS yielded no enriched pathways, prompting the threshold relaxation to 5×10^{-5} for this disease only.

2.4 GREAT analysis

The online tool GREAT associates both proximal and distal input genomic regions to their target gene, considering potential regulatory impact of distal genomic regions on gene function (McLean *et al.*, 2010; Tanigawa, Dyer and Bejerano, 2022). Datasets must be uploaded into GREAT in a BED file format. At minimum, BED files must contain columns for chromosome number, start position and end position. Datasets were reformatted into BED files and had columns ‘Chromosome’ (chromosome number), ‘Position_start’ (start location), ‘Position_end’ (start location +1), ‘A1’, ‘A2’, and ‘P’. The appropriate GRCh (Genome Reference Consortium human genome) was selected, according to the original study, GRCh37 for every disease except LBD which was GRCh38. The gene to genomic region association table obtained from the GREAT analysis was downloaded and reuploaded into R, prepared for pathway enrichment analysis.

2.4.1 Gene overlap

In RStudio the intersect function can be used to find the intersection between 2 objects, it was used to find the genes shared between disease pairs. The gene lists were also inputted into Venny (2.1.0) (Oliveros, 2015), an interactive tool for comparing lists with Venn diagrams. This tool was employed to calculate the percentage similarity between disease-associated gene lists for each pair of diseases in a 2-way comparison. Results were manually visualised as Venn diagrams created in Biorender.com.

2.5 Pathway enrichment

To gain insight into the underlying pathway mechanisms, pathway enrichment analysis was performed using genes obtained from GREAT analysis. For GO enrichment, the enrichGO function from the clusterProfiler package was used. This database was chosen for its standardised format that facilitates straightforward comparison between the diseases and downstream analysis. The function compares a set of input genes against a reference database to identify overrepresented GO terms and utilises gene-to-GO term annotations from the org.Hs.eg.db package to provide gene annotations. The GO Biological Processes (BP) ontology branch was selected for this analysis, capturing functional annotations related to biological processes. The analysis utilised the default parameters of the enrichGO function, including a p-value cut-off of 0.05, a q-value cut-off of 0.2 and the Benjamini-Hochberg method for multiple testing correction. Pathways with an adjusted P-value of <0.05 were considered significantly enriched.

2.5.1 Pathway overlap

The selected GOSemSim method, proposed by Wang et al (2007) calculates the semantic similarity between two GO terms based on their location in the GO graph and their relations with their ancestor terms. Generated scores range from 0-1, with 0 signifying no semantic similarity between the terms, while 1 indicates that the terms are identical. This method acknowledges the hierachal structure of GO terms and considers the position of the term to understand its semantic context. Additionally, it introduces the concept of S-values, which are a quantitative measure of how parent and child terms contribute to the semantics of a given term. Unlike other methods that depend on the frequencies of a set of GO terms and their closest common ancestor term, Wang's employment of graph-based methods gives a more comprehensive understanding on semantic similarity beyond simple measures. The method of calculation is as follows:

$$sim_{Wang}(A, B) = \frac{\sum_{t \in T_A \cap T_B} S_A(t) + S_B(t)}{SV(A) + SV(B)}$$

S-values ($S_A(t)$ and $S_B(t)$) quantify how much term ‘ t ’ contributes to the semantics of term A and B, respectively. Semantic values ($SV(A)$ and $SV(B)$) are the sum of all S-values in the hierarchy of A and B, respectively.

To get a single score, the Best Matched Average (BMA) strategy was selected, calculating the average of all maximum similarities across each row and column. It is defined as:

$$sim_{BMA}(g_1, g_2) = \frac{\sum_{1=i}^m \max_{1 \leq j \leq n} sim(go_{1i}, go_{2j}) + \sum_{1=j}^n \max_{1 \leq i \leq m} sim(go_{1i}, go_{2j})}{m + n}$$

BMA scores were calculated between the disease associated GO terms, comparing the diseases every 2-ways. Scores were then visualised in a Venn diagram, created in Biorender.com.

Further GOSemSim analysis used the uncombined parameter to get a matrix of scores between each individual GO term within one set of disease pathways and every individual GO term within another set of disease pathways, a process repeated for every possible disease pairing. GO terms with semantic similarity scores of > 0.8 were considered shared terms within the disease pairings. The number of overlapping GO terms with a semantic similarity score of > 0.8 between the diseases in 2-way, 3-way, 4-way, 5-way and 6-way comparison were reported, excluding those with no overlaps. The GO terms were ranked according to the number of diseases in which they appeared and annotated with its corresponding descriptive term using the org.Hs.eg.db package in RStudio.

3 Results

3.1 Overlaps in disease-associated genes

Filtering of SNPs by 5.0×10^{-8} for AD, PD, LBD, HD and MS, and 5.0×10^{-5} for ALS revealed significant disease variants for AD ($n = 3570$), PD ($n = 3465$), ALS ($n = 546$), LBD ($n = 189$), HD ($n = 230$) and MS ($n = 26,396$). Analysis of genes associated with AD ($n = 185$), PD ($n = 114$), ALS ($n = 223$), LBD ($n = 25$), HD ($n = 5$) and MS ($n = 412$) revealed no overlapping genes across all 6 diseases. However, the following genes were identified across 4 diseases (AD, PD, ALS and MS): *BTNL2*, *HLA-DQA1*, *HLA-DQA2* and *HLA-DRA* (Figure 14A). In addition, *HLA-DRB1* was shared across AD, PD and MS, *HLA-DBR5* was shared across AD, ALS, MS and *DGKQ* and *TMEM175* were shared across PD, ALS and LBD (Figure 14). All other gene overlaps were pairwise, in AD-PD ($n = 5$), AD-ALS ($n = 7$), AD-LBD ($n = 7$), AD-MS ($n = 7$), PD-ALS ($n = 9$), PD-LBD ($n = 13$), PD-MS ($n = 7$), ALS-LBD ($n = 2$) and ALS-MS ($n = 24$) (Figure 12; Figure 13). No overlaps were found between HD and any other disease, nor between LBD and MS.

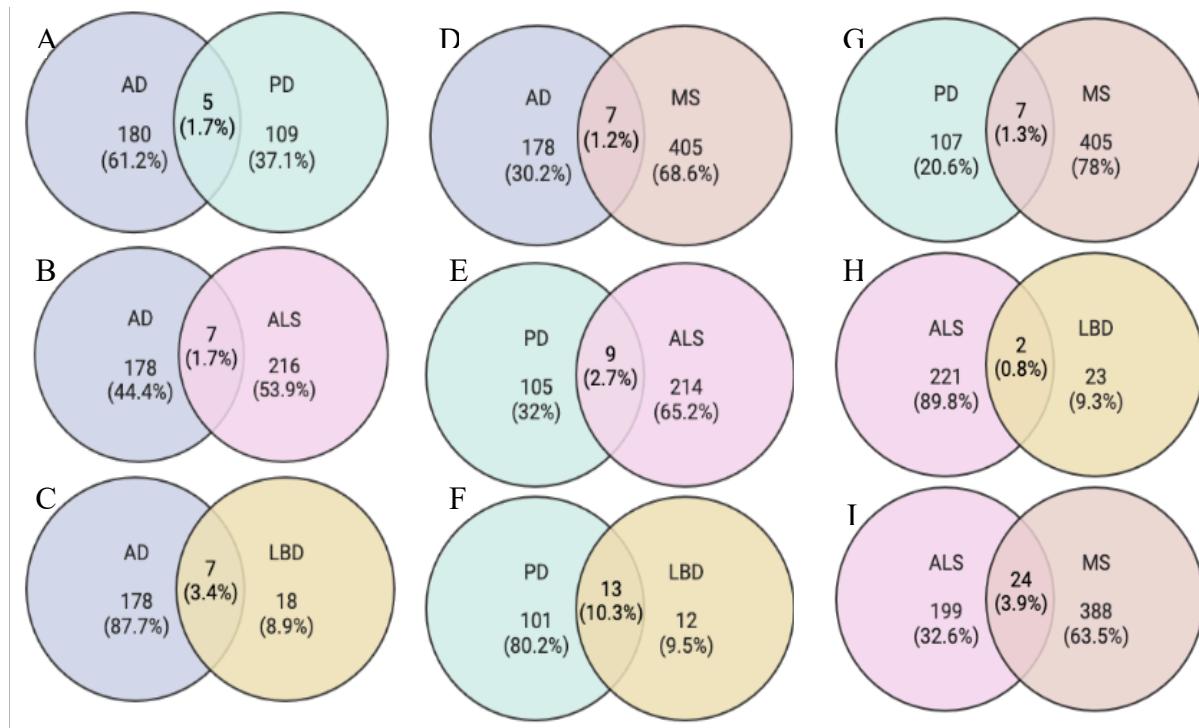


Figure 12- Pairwise overlapping genes across AD, PD, ALS, LBD and MS. Significant overlapping genes across the diseases pairwise, visualised with their percentage similarity calculated using online tool Venny (2.1.0). The number of overlapping genes and percentage similarity were: AD-PD ($n = 5$, 1.7%)(A), AD-ALS ($n = 7$, 1.7%)(B), AD-LBD ($n = 7$, 3.4%)(C), AD-MS ($n = 7$, 1.2%)(D), PD-ALS ($n = 9$, 2.7%)(E), PD-LBD ($n = 13$, 10.3%)(F), PD-MS ($n = 7$, 1.3%)(G), ALS-LBD ($n = 2$, 0.8%)(H) and ALS-MS ($n = 24$, 3.9%)(I). No overlaps were identified between HD and the other diseases or between LBD and MS. See Figure 2 for the full list of overlapping genes corresponding to the Venn diagrams in this image. Created in BioRender.com

A	B	C	D	I
AD & PD BTNL2 HLA-DQA1 HLA-DQA2 HLA-DRA HLA-DRB1	AD & ALS BTNL2 GPX3 HLA-DQA1 HLA-DQA2 HLA-DRA HLA-DRB5 TNIP1	AD & LBD APOC1 APOC4 APOE BCAM BIN1 CYP27C1 TOMM40	AD & MS BTNL2 C6orf10 HLA-DQA1 HLA-DQA2 HLA-DRA HLA-DRB1 HLA-DRB5 NOTCH4	ALS & MS AGPAT1 BAG6 BAK1 BTNL2 CDK4 CTDSP2 EGFL8 HIST1H2BJ HLA-B HLA-DOB HLA-DQA1 HLA-DQA2 HLA-DQB2 HLA-DRA HLA-DRB5 KIF5A PIP4K2C POU5F1 PPT2 PRRC2A PRRT1 TSPAN31 ZBTB9 ZNF322
E	F	G	H	
PD & ALS BTNL2 CPLX1 DGKQ GAK HLA-DQA1 HLA-DQA2 HLA-DRA TMEM175 VPS13C	PD & LBD ASH1L DGKQ DPM3 GBA KRTCAP2 LMNA MEX3A MMRN1 MTX1 RUSC1 SNCA TIGD2 TMEM175	PD & MS ARHGAP27 BTNL2 HLA-DQA1 HLA-DQA2 HLA-DRA HLA-DRB1 SPATA32	ALS & LBD DGKQ TMEM175	

Figure 13- Pairwise overlapping genes between AD, PD, ALS, LBD and MS. Significant genes overlapping between of AD-PD ($n = 5$)(A), AD-ALS ($n = 7$)(B), AD-LBD ($n = 7$)(C), AD-MS ($n = 7$)(D), PD-ALS ($n = 9$)(E), PD-LBD ($n = 13$)(F), PD-MS ($n = 7$)(G), ALS-LBD ($n = 2$)(H) and ALS-MS ($n = 24$)(I), corresponding to the Venn diagrams displayed in Figure 1. Created in BioRender.com

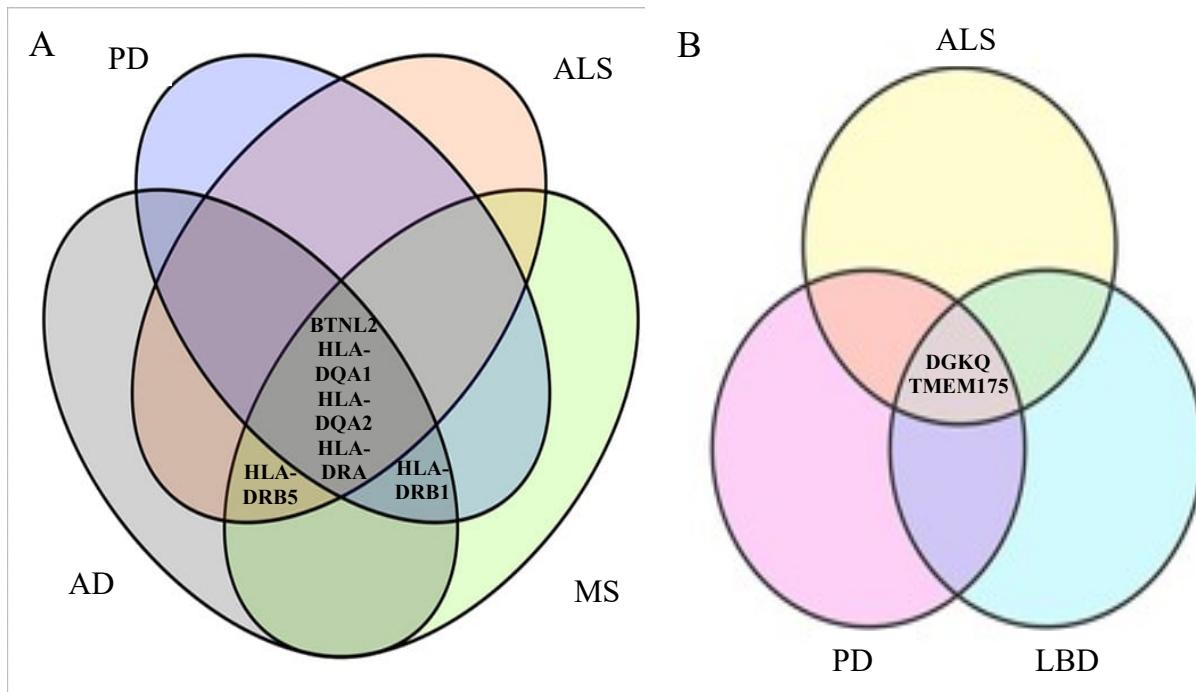


Figure 14- Overlap across more than 2 diseases. A- Showing the BTNL2, HLA-DQA1, HLA-DQA2 and HLA-DRA genes overlapping across AD, PD, ALS and MS, with HLA-DRB5 being exclusive to AD, ALS and MS and HLA-DRB1 to AD, PD and MS. B- Showing the DGKQ and TMEM175 genes shared across AD, PD and LBD.

3.2 Overlaps in key dysfunctional pathways

Pathway enrichment analysis of the key disease-associated genes via enrichGO revealed multiple shared dysfunctional pathways, in AD ($n = 222$), PD ($n = 48$), ALS ($n = 25$), LBD ($n = 78$), HD ($n = 56$), and MS ($n = 418$). Intersecting GO terms between AD-PD ($n = 19$), AD-ALS ($n = 12$), AD-LBD ($n = 30$), AD-HD ($n = 1$), AD-MS ($n = 96$), PD-ALS ($n = 14$), PD-LBD ($n = 11$), PD-MS ($n = 18$), ALS-DLB ($n = 3$), ALS-MS ($n = 12$), DLB-HD ($n = 2$), DLB-MS ($n = 2$), HD-MS ($n = 4$) were found. No overlaps were found between PD-HD and ALS-HD.

GOSemSim provided a BMA semantic similarity score of GO terms between diseases pairwise (Figure 15). The highest semantic similarity was between AD and MS with a score of 0.674, followed closely by PD and LBD with a score of 0.61. The disease pairings, ALS-HD and PD-HD had minimal semantic similarity (0.224 and 0.273, respectively); these were the same pairings in which no intersecting pathways were found between them. Similarly, a low semantic similarity score of 0.29 was calculated between ALS and MS, despite the 12 intersecting pathways identified between them.

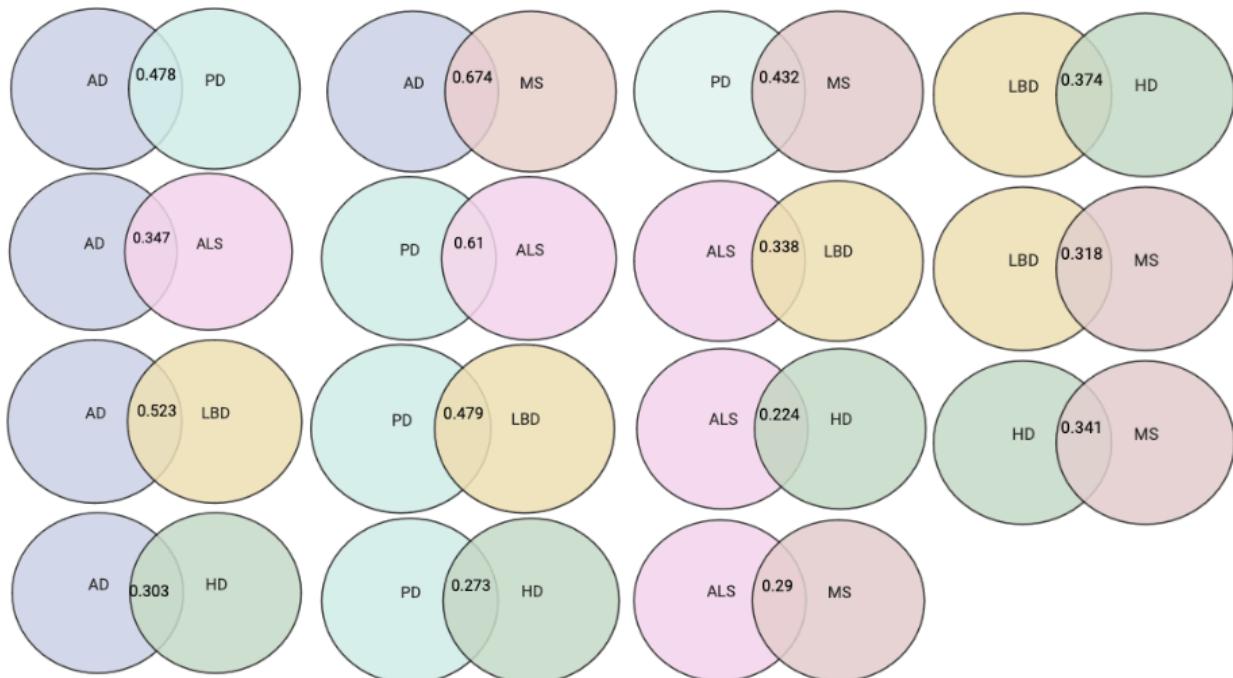


Figure 15- Semantic similarity scores across the diseases pairwise. BMA semantic similarity scores showing similarity between two sets of GO terms between the diseases pairwise, generated via R package GOSemSim. Scored from 0-1 with 0 being no semantic similarity and 1 denoting complete identity; AD-MS and PD-ALS showed a relatively high semantic similarity of 0.674 and 0.61, respectively. Conversely, ALS-HD, PD-HD and ALS-MS displayed relatively low scores with their respective semantic similarity scores being below 0.3. Created in Biorender.com

Further investigation explored the overlap between the diseases pairwise using a semantic similarity score threshold > 0.8 to assess each GO term for overlap. The results present the diseases combined every 2-ways and the corresponding number of GO terms present with a score of > 0.8 , ordered from the highest number of overlapping pathways to lowest (Table 2). The analysis revealed varying degrees of pathway overlap between disease pairs. The most substantial pathway overlap was evident in AD-MS with 141 overlapping GO terms identified. This was followed by AD-LBD ($n = 65$), AD-PD ($n = 31$) and PD-MS ($n = 27$). In the mid-range, the PD-ALS, AD-ALS, PD-LBD, and ALS-MS pairs exhibited pathway overlaps ranging from 13 to 15. Conversely the pathways showing fewer overlaps included HD-MS, LBD-MS, LBD-HD, ALS-LBD and AD-HD, each yielding 1-10 overlapping GO terms. Notably, PD-HD and ALS-HD did not generate any overlapping GO terms and were consequently excluded from Table 2.

Table 2- Pathway overlap between diseases every 2-ways. Each disease compared pairwise, with its corresponding number of overlapping GO terms with a semantic similarity score > 0.8 , ranked from the highest number of overlapping pathways to lowest. The greatest number of GO term overlaps can be observed between AD-MS ($n = 141$). PD-HD and PD-ALS were excluded as there were no overlapping GO terms. GOSemSim was used in R to assess semantic similarity.

Overlap between diseases every 2-ways	Number of overlapping GO terms with a semantic similarity score > 0.8 ranked from highest to lowest
AD-MS	141
AD-LBD	65
AD-PD	31
PD-MS	27
PD-ALS	15
AD-ALS	13
PD-LBD	13
ALS-MS	13
HD-MS	8
LBD-MS	7
LBD-HD	6
ALS-LBD	4
AD-HD	2

Subsequently, exploration to determine the overlap between diseases in an every 3-way combination was conducted, considering only the GO terms with a semantic similarity score > 0.8. Results are ranked from the highest number of overlapping pathways to lowest (Table 3). The AD-PD-MS trio displayed the highest degree of pathway overlap (n = 25), followed by AD-ALS-MS (n = 13), AD-PD-ALS (n = 12), and PD-ALS-MS (n = 11). Further examination revealed that among the AD-ALS-MS, AD-PD-ALS, and PD-ALS-MS trios, there were 11 shared GO terms. Notably, the term ‘antigen processing and presentation’ was exclusively shared between AD-ALS-MS and AD-PD-ALS, while ‘immunoglobulin mediated immune response’ was specific to the AD-ALS-MS overlap. In addition, all the GO terms found across the AD-ALS-MS, AD-PD-ALS, and PD-ALS-MS disease trios were also found within the AD-PD-MS pathway overlap, leaving 13 GO terms exclusive to the AD-PD-MS trio. On the other hand, PD-ALS-LBD, AD-PD-LBD, LBD-HD-MS, and AD-HD-MS each showed less pronounced overlap, each possessing fewer than 5 shared pathways within their respective trio. Every 3-way disease comparison with 0 overlapping pathways were not reported in Table 3.

Table 3- Pathway overlap between diseases every 3 ways. Each disease compared every 3 ways, with its corresponding number of overlapping GO terms with a semantic similarity score > 0.8, ranked from the highest number of overlapping pathways to lowest. The greatest number of overlaps can be seen in the disease trio AD-PD-MS, with 25 shared pathways. GOSemSim was used in R to assess semantic similarity.

Overlap between diseases every 3-ways	Number of GO terms with a semantic similarity score > 0.8 ranked from highest to lowest
AD-PD-MS	25
AD-ALS-MS	13
AD-PD-ALS	12
PD-ALS-MS	11
PD-ALS-LBD	4
AD-PD-LBD	3
LBD-HD-MS	1
AD-HD-MS	1

Finally, the 4-way exploration of overlap between the diseases revealed only one disease quartet. Only AD-PD-ALS-MS had overlapping GO terms with a semantic similarity score > 0.8 among them (Table 3; Table 4). This is the greatest number of diseases with overlapping pathways in this study. A 5-way and 6-way analysis failed to produce any pathway overlaps and were therefore not reported.

Table 4- Pathway overlap between diseases every 4-ways. Each disease compared every 4-ways, with its corresponding number of overlapping GO terms with a semantic similarity score > 0.8, ranked from the highest number of overlapping pathways to lowest. The only 4-way overlap was between AD-PD-ALS-MS with 12 shared GO terms between them. Other 4-way comparisons lacked overlapping pathways and were not reported. GOSemSim was used in R to assess semantic similarity

Overlap between diseases every 4-ways	Number of GO terms with a semantic similarity score > 0.8 ranked from highest to lowest
AD-PD-ALS-MS	12

The top pathways across the diseases were selected based on the number of diseases in which the GO term appears (Table 5). Among these, the top 12 pathways appear across AD, PD, ALS and MS. These pathways are MHC and MHC II protein complex assembly, peptide antigen assembly with MHC and MHC II, antigen processing and presentation, and immunoglobulin production in an immunoglobulin-mediated immune response. Across AD, PD, and MS, there are 13 pathways exclusive to this subset. These pathways include the differentiation of various types of T-cells, such as T-helper cells and T-memory cells. Additionally, they encompass the differentiation of different T-cell types in an immune response, including CD4 positive alpha-beta T-cells and alpha-beta T-cells. Moreover, the pathways also cover the activation of T-cells involved in an immune response, as well as alpha-beta T-cell activation. Furthermore, the list includes pathways related to lymphocyte activation, immune responses mediated by immunoglobulin, and positive regulation of cell, lymphocyte, and leukocyte activation. A total of 4 pathways are specific to PD, ALS and LBD, these relate to the synaptic vesicle cycle, its recycling, transport in the synapse and assembly. The 3 pathways shared across AD, PD and LBD are related to endocytosis and its regulation, more specifically receptor-mediated endocytosis. The three pathways shared across AD, PD, and LBD involve endocytosis, its regulation (including positive regulation),

and receptor-mediated endocytosis. Only a single shared pathway was identified between LBD, HD and MS: regulation of monooxygenase activity. Similarly, a single term was shared between AD, HD and MS: somatic cell DNA recombination.

Table 5- The top pathways across the diseases based on the number of diseases in which the term appears. The columns show the GO term ID, its corresponding term and a colour pattern of the diseases in which the term is present. recombination was seen across PD, HD and MS. GO terms were enriched by R package org.Hs.eg.db (3.17.0).

GO ID	Term	Key:	
GO:0002399	MHC class II protein complex assembly	AD PD ALS LBD HD MS	
GO:0002396	MHC protein complex assembly		
GO:0002503	peptide antigen assembly with MHC class II protein complex		
GO:0002501	peptide antigen assembly with MHC protein complex		
GO:0002495	antigen processing and presentation of peptide antigen via MHC class II		
GO:0019886	antigen processing and presentation of exogenous peptide antigen via MHC class II		
GO:0002478	antigen processing and presentation of exogenous peptide antigen		
GO:0002504	antigen processing and presentation of peptide or polysaccharide antigen via MHC class II		
GO:0048002	antigen processing and presentation of peptide antigen		
GO:0019882	antigen processing and presentation		
GO:0019884	antigen processing and presentation of exogenous antigen		
GO:0002381	immunoglobulin production involved in immunoglobulin-mediated immune response		
GO:0002292	T cell differentiation involved in immune response		
GO:0042093	T-helper cell differentiation		
GO:0002294	CD4-positive, alpha-beta T cell differentiation involved in immune response		
GO:0002293	alpha-beta T cell differentiation involved in immune response		
GO:0002285	lymphocyte activation involved in immune response		
GO:0002286	T cell activation involved in immune response		
GO:0002287	alpha-beta T cell activation involved in immune response		
GO:0016064	immunoglobulin mediated immune response		
GO:0043379	memory T cell differentiation		
GO:0030217	T cell differentiation		
GO:0002696	positive regulation of leukocyte activation		
GO:0051251	positive regulation of lymphocyte activation		
GO:0050867	positive regulation of cell activation		
GO:0036465	synaptic vesicle recycling		
GO:0099504	synaptic vesicle cycle		
GO:0099003	vesicle-mediated transport in synapse		
GO:0016050	vesicle organization		
GO:0006898	receptor-mediated endocytosis		
GO:0030100	regulation of endocytosis		
GO:0045807	positive regulation of endocytosis		
GO:0032768	regulation of monooxygenase activity		
GO:0016444	somatic cell DNA recombination		

4 Discussion

4.1 Overview

A deeper understanding of molecular and cellular mechanisms underlying NDDs is essential for the therapeutic development required to maintain the health and functionality of the steadily increasing elderly population. GWAS have revealed novel loci and NDD-associated genes, however despite this deeper insight, it has not resulted in a cure. By focusing on the commonalities between NDDs, specifically their shared underlying genes and pathways, a deeper insight can be obtained into which specific mechanisms should be targeted in treatment. Targeting pathways found across multiple diseases could cure, delay onset, slow progression or at least alleviate symptoms of all the diseases the pathway is implicated in. The project aimed to use various methods to assess the shared underlying genes and mechanisms across AD, PD, ALS, LBD, HD and MS.

4.2 Gene and pathway overlap

The following genes were found to be significantly enriched via GREAT analysis across AD, PD, ALS and MS: BTNL2, HLA-DQA1, HLA-DQA2 and HLA-DRA (Figure 14). BTNL2, also known as butyrophilin-like 2 is a gene belonging to the immunoglobulin gene superfamily, it encodes a major histocompatibility complex (MHC) class II association (Konno *et al.*, 2009; Sayers *et al.*, 2022). It serves as a negative T-cell regulator by binding to a putative receptor on activated T-cells to inhibit its proliferation and cytokine release (Nguyen *et al.*, 2006). This gene resides within the HLA region and its mutation has been significantly associated with inflammatory diseases. This class II molecule is expressed in antigen presenting cells (APCs; including B-lymphocytes, dendritic cells and macrophages) and plays a role in the immune system by presenting peptides derived from extracellular proteins (Sayers *et al.*, 2022). HLA-DQA2 also belongs to the class II alpha chain family, the encoded protein forms a heterodimer with a class II beta chain (Stelzer *et al.*, 2016). The protein is located inside intracellular vesicles and plays a role in the peptide loading of MHC class II molecules, it binds peptides from antigens that enter APCs via endocytosis, presenting them on the cell surface for recognition by CD4 T-helper cells (Lenormand *et al.*,

2012). The molecule is expressed on various APCs and plays a central role in the immune system by presenting peptides from extracellular proteins, specifically from pathogens to CD4 T-helper cells to initiate an immune response (Sayers *et al.*, 2022).

HLA genes often form heterodimers with one another, a process in which two protein chains come together to form an MHC complex. Alpha and beta chains come together to form a functional MHC class II molecule which can then be expressed on the surface of APCs (Sayers *et al.*, 2022). Dimer formation between HLA-DQB2 and HLA-DQA2 plays an important role in immunology, with them having to associate with the invariant chain to reach endosomal compartments whilst HLA-DQB1 and HLA-DQA2 formed mixed heterodimers that efficiently left the endoplasmic reticulum (Lenormand *et al.*, 2012).

The clear commonality across these genes is that they are all MHC genes, the human leukocyte antigen (HLA) is the human version of the MHC. The function of the MHC is to bind peptide fragments from pathogens and foreign matter and display them on the cell surface for recognition by the appropriate T-cells. MHC class I molecules are expressed by nucleated cells for immune surveillance; however, it is the MHC class II genes that have been significantly associated with AD, PD, ALS and MS in this study (Figure 14). MHC class II complexes are expressed on the cell surface of APCs, that present epitopic peptides to CD4⁺ T cells and initiate adaptive immune responses including secretion of cytokines and generation of antibodies (Mangalam, Taneja and David, 2013). The MHC is polygenic and highly polymorphic. This diversity allows for a greater range of MHC molecules, which in turn enables the recognition of a wider variety of antigens and the generation of an immune response when required (Charles A Janeway *et al.*, 2001).

The AD, PD and MS subset identified the HLA-DRB1 genes as disease-associated, while the AD, ALS and MS subset identified the HLA-DRB5 gene as disease-associated (Figure 14). Both HLA genes have similar roles in the initiation of an immune response. However, unlike those mentioned above, they belong to the HLA class II beta chain paralogs. Both molecules form a heterodimer with HLA-DRA. Within their respective DR molecule, they contain all the polymorphisms that determine peptide binding.

GWAS have identified a single SNP within the *HLA-DRB1* as a risk factor for LOAD (Lambert *et al.*, 2013). One study suggested the genetic effect of *HLA-DRB1* is predisposing

factor to LOAD in the absence of the *APOE* ε4 allele, while *APOE* ε4 allele is the most susceptible genetic factor in *APOE* ε4 carriers (Lu *et al.*, 2017). HLA-DRB5 was also identified as a novel locus significantly associated to the risk of AD in Caucasian and Chinese population (Lambert *et al.*, 2013; Lu *et al.*, 2017). The presence of HLA-DRB1 and HLA-DRB5 expression in microglia within the prefrontal cortex of 48 individuals exhibiting varying degrees of AD pathology showed a positive correlation with measures of AD pathology (Mathys *et al.*, 2019). A study fine mapping the HLA locus to AD determined the strongest association with AD risk with specific MHC haplotypes in the DRB1, DQA1 and DQB1 HLA genes, with secondary analysis suggesting this effect may be driven by individuals negative for the *APOE* gene (Steele *et al.*, 2017). It is now generally accepted that upregulation of HLA class II antigens are a definitive marker of activated microglia which have been shown to be involved in the formation of lesions in AD (McGeer, McGeer and Yasojima, 2000).

Associations between SNPs within the HLA-DRA region and late-onset sporadic PD have been observed (Hamza *et al.*, 2010). These findings have been independently replicated, and subsequent studies have extended these observations, to include the *HLA-DRB5* and *HLA-DRB1* gene loci (Ahmed *et al.*, 2012), even though HLA-DRB5 was not significantly associated with PD in this study (Figure 14). There is evidence to suggest that HLA-DRA PD-associated polymorphisms, may have an increased risk by higher baseline and induced levels of MHC class II proteins on blood monocytes, B-cells, and a shift in their phenotype towards promoting inflammatory responses (Kannarkat *et al.*, 2015). It should be noted that HLA-DRA shows a low level of polymorphisms and is not prominent in association studies, therefore attempts to replicate findings of an association have yielded contradictory results (Guo *et al.*, 2011; Do *et al.*, 2011).

Very few genetic association studies have evaluated the HLA region in ALS. Initial studies found no correlation between HLA antigens and ALS in patients (Bartfeld *et al.*, 1982). Some later studies linked HLA class I genes to increased risk of ALS, but these findings were regarded inconsistent in the identifying a link between a particular HLA antigen and ALS in study populations (Misra, Damotte and Hollenbach, 2018). However, a recent study indicated a role of HLA class II molecules in ALS, more specifically polymorphisms in the HLA-DRA and HLA-DRB5 genes were linked to increased risk of sporadic ALS (Yang *et al.*, 2017). It should be noted, this study was limited to a southwest Chinese cohort, but while

inconclusive, it suggests an immunogenetic component to ALS, warranting further examination.

1972 was the first association of HLA class I antigens with MS (Naito *et al.*, 1972). It later became apparent these alleles were a part of an extended class I and II haplotype (Hauser *et al.*, 1989). A report reviewing 72 studies from 1993 to 2004 observed significantly increased frequency of the receptor DRB1*15:01 among most MS patients (Schmidt, Williamson and Ashley-Koch, 2007). A GWAS similarly confirmed that the main susceptibility for MS maps to DRB1 in the class II HLA region, describing up to 10.5% of genetic variance underlying MS risk (International Multiple Sclerosis Genetics Consortium *et al.*, 2011). HLA-genotyping analysis revealed a robust correlation between specific HLA-DQB1 alleles and increased severity in inflammation and neurodegeneration observed in MRI scans (Zivadinov *et al.*, 2007).

The HLA genes observed across AD, PD ALS and MS (Figure 14), correlate to the pathways related to MHC class II protein assembly, peptide antigen assembly, antigen processing and presentation and T-cell differentiation and activation (Table 5). An investigation on HLA mechanisms on AD pathogenesis determined that A β peptides can be swallowed by microglia, processed into peptides and presented to T-lymphocytes after combination with a HLA molecule. This process can stimulate B lymphocytes to secrete antibodies and T-lymphocytes to kill cells producing excess A β (Wang, Wan and Xing, 2020). However, this reaction in excess can be very harmful (Malm *et al.*, 2008). Another piece of evidence of immune response mediated inflammation in AD is that nonsteroidal anti-inflammatory drugs were found to be an effective AD treatment, likely by inhibiting microglial activation, implying that HLA may act in inflammatory responses in AD (Mackenzie and Munoz, 1998). Therefore, immune system mediated inflammation can perhaps explain the association between AD risk and HLA molecules. To obtain a deeper understanding on the specific pathways affected that contribute to AD pathogenesis, further study is required. An A β vaccination was proposed as a therapeutic option, demonstrating A β clearance and inhibition of A β -mediated glial activation in animal models (Monsonego *et al.*, 2003; Orgogozo *et al.*, 2003). However, A β -reactive T-cell responses showed varying responses bases on MHC haplotypes. For example, T-cell proliferation could be blocked by anti-HLA-DR antibodies but not HLA-DQ or HLA-DP blockers, therefore making it an unsuitable treatment option for

individuals whose T-cell responses to A β are influenced by HLA-DQ and HLA-DP (Monsonego *et al.*, 2003).

A study identifying HLA profiles in PD and dementia determined PD has a weak correlation with susceptibility alleles compared to dementia (James and Georgopoulos, 2021). HLA could be involved in protection from or susceptibility to these diseases. HLA class II alleles play a critical role in facilitating antibody production and immunological memory. Highly protective alleles effectively promote pathogen elimination that may have otherwise contributed to disease susceptibility (James and Georgopoulos, 2020). In the absence of HLA protection against pathogens, disease may result from directly damaging effects of a pathogen on cells or because of susceptibility HLA alleles that promote autoimmunity due to chronic inflammation (James and Georgopoulos, 2021).

In ALS, evidence supports a role for autoimmune mechanisms in motor neuron degeneration, IgG antibodies have been detected in the upper and lower motor neurons, and T-lymphocytes and activated microglia have been identified within spinal cord grey matter and motor cortex, therefore implying a role of autoimmune mechanisms in the loss of motor neurons in ALS (Appel *et al.*, 1994). Similar findings indicated neuroinflammation in ALS, MHC class II glycoproteins become upregulated when immune cells were activated (McGeer and McGeer, 2002). When studied using immunohistochemistry, the appearance of activated microglia was observed (McGeer and McGeer, 2002). Inflammation mediated by the immune system can cause oxidative stress, generating ROS species further exacerbating cellular damage. COX-inhibitors may be an effective therapeutic option in treating ALS due to its high upregulation in ALS and association with inflammatory pathways (McGeer and McGeer, 2002). However, observed manifestations of inflammation and oxidative stress are only one component of more complex inflammatory response, showing the need to further uncover underlying mechanisms.

The HLA association is consistent with the idea that MS is an antigen-specific autoimmune disease driven by several mechanisms. In both white and grey brain matter in the CNS, inflammation caused by immune cells and their cytokines infiltrating tissue is an early cause of damage in MS (Ghasemi, Razavi and Nikzad, 2017). In addition, pathogen-associated molecules bind to toll-like receptors on APCs, producing cytokines including IL-12, IL-23, and IL-4. These can induce CD4 T-cell differentiation into Th1, Th2, or Th17 phenotypes that

can release cytokines, including proinflammatory cytokines like IFN γ , TNF- α , IL-17, IL-21, IL-22, and IL-26 and anti-inflammatory cytokines such as IL-4 and IL-13 that contribute to MS progression (Schoenborn and Wilson, 2007; Zhu and Paul, 2008; Minty *et al.*, 1993). Furthermore, CD8 T-cells have been identified in MS lesions (Kouchaki *et al.*, 2014), these cells increase vascular permeability facilitating immune cell infiltration, destroying glial cells and triggering oligodendrocyte death. In addition to CNS inflammation, the myelin repair process due to oligodendrocyte death is also impaired (Kasper and Shoemaker, 2010), making repair during recovery periods difficult, explaining why MS can lead to accumulated damage over time. Given the disease heterogeneity, there is no single therapeutic target for MS, the main goal of therapeutic treatment is to quiet the disease by reducing inflammation, myelin injury and relapses (Yang *et al.*, 2022).

The DGKQ gene that encodes the diacylglycerol kinase theta protein (DGK θ) was identified as disease-associated across PD, ALS and LBD (Figure 14). It is localised in the speckle domains of the nucleus and converts diacylglycerol (DAG) into phosphatidic acid (PA), bioactive lipid molecules that play a role as second messengers in intracellular signalling pathways (Schoenborn and Wilson, 2007). DGK θ converts DAG into PA by phosphorylating it, regulating the balance between DAG and PA. DAG supports the biosynthesis and degradation of glycerolipids and regulates protein kinase C (PKC) activity (Bishop and Bell, 1988), which influences various processes like regulation of neurotransmitter release, ion channels, growth and differentiation, and neural plasticity (Huang, 1989). PA is involved in processes related to membrane curvature, vesicle trafficking and regulation of protein kinases (Stillwell, 2016). By DGK θ regulating the balance between DAG and PA, it modulates signalling pathways, allowing the maintenance of proper cellular response. This regulation is particularly important in the nervous system where it is now well-established that DAG and PA serve important roles in many neurological functions (Tu-Sekine and Raben, 2011).

The pathways enriched in PD, ALS and LBD are those related to the synaptic vesicle (SV) cycle, SV recycling, vesicle-mediated transport in the synapse and vesicle organisation (Table 5). The DAG and PA molecules regulated by DGKQ have been implicated in synaptic function. DAG may play 3 roles in the SV cycle (Goldschmidt *et al.*, 2016). Firstly, DAG enhances Munc13-1 activity, vital for efficient priming of SVs before release during synaptic transmission (Bauer *et al.*, 2007). Secondly, it activates PKC that phosphorylates and thereby regulates activity of presynaptic cell proteins, more specifically SNARE complex protein

including Munc-18 and SNAP-25 (Paolo *et al.*, 2004; Rhee *et al.*, 2002), key players in SV fusion and neurotransmitter release (Tang *et al.*, 2021). Lastly, termination of DAG by DGK θ via phosphorylation generates PA, also involved in vesicle trafficking, cell signalling and neurotransmitter release (Antonescu, Danuser and Schmid, 2010). Furthermore, knockdown of DGK θ significantly slowed the rate of SV endocytosis compared to control neurons (Goldschmidt *et al.*, 2016).

Specific genetic variants in the *DGKQ* region have been repeatedly reported to be associated with PD (Pankratz *et al.*, 2009; Simón-Sánchez *et al.*, 2011), however little is known about the function of *DGKQ*. One possible explanation put forward is that the specific variants can affect splicing sites potentially leading to altered protein products that may be involved in the pathogenesis of PD (Chen *et al.*, 2013). DGKQ was listed as a candidate drug target for PD, but the mechanisms and clinical utility warrant further exploration by experimental studies (Ge *et al.*, 2023). There is a lack of reported associations between ALS and DGKQ and LBD and DGKQ. The exact role of DGKQ in these NDDs is not well understood, even with some connections being established in PD, further research is needed to explore the impact of this gene. In this study DGKQ was also identified in ALS and LBD (Figure 14), although this correlation has not been widely reported in literature. This discovery needs confirmation from additional studies, and if validated could be a potential target for drug development.

TMEM175 (transmembrane Protein 175) was identified as a disease-associated gene in PD, ALS and LBD (Figure 14), this aligned with the results of a recent study (Wightman *et al.*, 2023). This protein is located in cellular membranes of endosomes and lysosomes. It functions as a proton-activated, highly selective proton channel that catalyses the release of protons from endosomes and lysosomes to maintain a steady luminal pH. Activated at low pH levels by luminal side protons: it mediates the release of protons from lysosomes, prompting a proton leak that balances activity of V-ATPase, an enzyme responsible for regulating pH within lysosomes (Simón-Sánchez *et al.*, 2011). An abnormal lysosomal pH impairs lysosomal degradation, cargo loading, catabolite export, vesicle movement, and nutrient sensing (Ballabio and Bonifacino, 2020; Xu and Ren, 2015), contributing to pathologies in PD.

It was observed that lysosomes lacking TMEM175 exhibit no potassium conductance, with little sensitivity to altered potassium levels, therefore compromising lysosomal pH stability

and autophagosome-lysosome fusion (Xu and Ren, 2015). However, it was later concluded that under normal physiological conditions, proton conductance, not potassium conductance was primarily responsible for the pH-dependent lysosomal functions reported (Hu *et al.*, 2022). Researchers determined that TMEM175 loss of function resulted in lysosomal hyper-acidification and impaired proteolytic activity of lysosomal hydrolases. Notably after pH was optimised, lysosomal activity was restored. Furthermore, TMEM175-knockout mice showed pathological aggregation of α -synuclein in the brain, lysosomal over-acidification and impaired lysosomal hydrolytic activity (Hu *et al.*, 2022). Thus, identifying the modulation of lysosomal pH as a potential therapeutic target for improving lysosomal degradation most seen in PD, AD, and other pathologies associated with dysregulated lysosome function (Perdigoto, 2022).

For efficient degradation of macromolecules, lysosomes require an acidic lumen, disruptions to lysosomal pH can impair lysosomal function, representing a basis for pathogenic protein accumulation linked to pathologies like PD and AD (Perdigoto, 2022; Lee *et al.*, 2010). It is likely, ineffective lysosomal function results in the accumulation of protein aggregates as they cannot undergo degradation via autophagy, causing subsequent cellular damage. While the literature associated TMEM175 predominantly with PD, when considering the shared underlying α -synuclein pathology with LBD, it is logical for the gene to be implicated in both NDDs in this study (Figure 14). Additionally, the literature linked this gene to AD and its pathology, but this was not determined in this study.

Despite the lack of initial reports on the association between TMEM175 and ALS, it's worth noting that disruptions in autophagy and endolysosomal pathways have consistently been reported in ALS (Sullivan *et al.*, 2017; Bordoni *et al.*, 2022). The identification of this new loci in ALS prompts the need for additional research to validate and understand the significance of the observed association. It is possible this gene may be connected to the lysosomal dysfunction already observed in ALS, potentially marking a novel locus that could be targeting for therapeutic interventions.

One study has already proposed a potential therapeutic measure targeting dysregulated lysosomal pathways in ALS. It was observed that treatment with Trehalose, a disaccharide that modulates autophagy, decreased LC3 levels (a protein associated with autophagy), likely inducing the degradation of accumulated autophagolysosomes typically observed in ALS

(Bordoni *et al.*, 2022). Given that these dysfunctional pathways are also observed across AD, PD and LBD, Trehalose's potential therapeutic effects may extend to these diseases, however further research would be required to validate its efficacy and safety.

4.3 Limitations and future work

There are several limitations to this study. Firstly, the summary statistics acquired while were from the most comprehensive GWAS, were all limited to the European ancestry thus making findings ungeneralisable to other populations across the globe because genetic variants differ among ethnic groups. In addition, research focused on only one population pre-dominantly benefits that group. For example, therapeutic developments targeting variants, genes and pathways identified in only one cultural group may be ineffective among other ethnicities, raising ethical concerns. Given this, in future work, researchers can expand their variety of participants to include a more diverse range of individuals, leading to more generalisable data across global populations. In addition, validation studies across different populations are crucial to determine if previous findings align and the extent to which previously identified genetic associations hold true.

Another limitation is the sole use of genomic data in this study to determine the underlying pathways and mechanisms in disease. The initial study was supposed to be a collaboration of genomic and transcriptomic data, providing additional insight into gene expression levels revealing active genes. As not all transcribed genes directly translate into functional protein products, proteomic data could have also been employed to understand gene function more directly. This, however, was unfortunately not possible. The design of the study had to be updated to only include genomic data from GWAS due to time constraints. In addition, the recruitment of proteomic data, would have been very difficult due to the lack of existing data. As different data types all come with their own benefits and limitations, future studies can strive to integrate layers of data such as genomics, transcriptomics and proteomics. This approach would provide a more in-depth analysis of gene expression, regulation, function, enhancing the accuracy of identifying underlying disease mechanisms, pathways and pathologies.

The study faced several other limitations that should be considered. Firstly, the inability to access the complete datasets for each disease due to data protection laws, may have resulted in a less in-depth exploration of associations. Secondly, the study's use of GWAS that focuses on genetic variants with a small to moderate effect size, potentially disregards rare variants with substantial impact. This arises due to the infrequent occurrence of these variants in individuals. Lastly, the decision to reduce the significance threshold exclusively for ALS, while was necessary due to the lack of outcomes when adhered to the standard threshold (5×10^{-8}), could have introduced a bias. Fortunately, the results were reassuring as they did not appear skewed, evidenced by the consistent overlap with associations identified in literature.

4.4 Conclusion

In summary, this study identified overlapping HLA genes across AD, PD, ALS and MS and their corresponding pathways related to MHC class II molecules, antigen processing and presentation, T-cell differentiation and activation. Overlapping genes identified in PD, ALS and LBD related to the SV cycle, SV recycling, vesicle-mediated transport in the synapse, vesicle organisation, and endocytosis and its regulation. These findings strongly suggest the dysfunction of the immune system as a key feature across these NDDs, prompting the need to conduct further studies to explore these associations and potentially identify more specific mechanisms that can be targeted for therapeutic interventions.

5 References

- Ahmed, I. *et al.* (2012) ‘Association between Parkinson’s disease and the HLA-DRB1 locus’, *Movement Disorders: Official Journal of the Movement Disorder Society*, 27(9), pp. 1104–1110. Available at: <https://doi.org/10.1002/mds.25035>.
- Albin, R.L. *et al.* (1990) ‘Striatal and nigral neuron subpopulations in rigid Huntington’s disease: Implications for the functional anatomy of chorea and rigidity-akinesia’, *Annals of Neurology*, 27(4), pp. 357–365. Available at: <https://doi.org/10.1002/ana.410270403>.
- An, H. *et al.* (2019) ‘ALS-linked FUS mutations confer loss and gain of function in the nucleus by promoting excessive formation of dysfunctional paraspeckles’, *Acta Neuropathologica Communications*, 7(1), p. 7. Available at: <https://doi.org/10.1186/s40478-019-0658-x>.
- Antonescu, C.N., Danuser, G. and Schmid, S.L. (2010) ‘Phosphatidic acid plays a regulatory role in clathrin-mediated endocytosis’, *Molecular Biology of the Cell*, 21(16), pp. 2944–2952. Available at: <https://doi.org/10.1091/mbc.E10-05-0421>.
- Appel, S.H. *et al.* (1994) ‘Evidence for autoimmunity in amyotrophic lateral sclerosis’, *Journal of the Neurological Sciences*, 124, pp. 14–19. Available at: [https://doi.org/10.1016/0022-510X\(94\)90171-6](https://doi.org/10.1016/0022-510X(94)90171-6).
- Ballabio, A. and Bonifacino, J.S. (2020) ‘Lysosomes as dynamic regulators of cell and organismal homeostasis’, *Nature Reviews Molecular Cell Biology*, 21(2), pp. 101–118. Available at: <https://doi.org/10.1038/s41580-019-0185-4>.
- Bal-Price, A., Moneer, Z. and Brown, G.C. (2002) ‘Nitric oxide induces rapid, calcium-dependent release of vesicular glutamate and ATP from cultured rat astrocytes’, *Glia*, 40(3), pp. 312–323. Available at: <https://doi.org/10.1002/glia.10124>.
- Bartfeld, H. *et al.* (1982) ‘HLA Frequencies in Amyotrophic Lateral Sclerosis’, *Archives of Neurology*, 39(5), pp. 270–271. Available at: <https://doi.org/10.1001/archneur.1982.00510170012003>.
- Batino, L.K.J. *et al.* (2021) ‘Sporadic Huntington’s disease in the Philippines: a case report’, *Neurodegenerative Disease Management*, 11(6), pp. 445–449. Available at: <https://doi.org/10.2217/nmt-2021-0023>.
- Bauer, C.S. *et al.* (2007) ‘Potentiation of exocytosis by phospholipase C-coupled G-protein-coupled receptors requires the priming protein Munc13-1’, *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 27(1), pp. 212–219. Available at: <https://doi.org/10.1523/JNEUROSCI.4201-06.2007>.
- Bishop, W.R. and Bell, R.M. (1988) ‘Functions of diacylglycerol in glycerolipid metabolism, signal transduction and cellular transformation’, *Oncogene Research*, 2(3), pp. 205–218.
- Blennow, K., Leon, M.J. de and Zetterberg, H. (2006) ‘Alzheimer’s disease’, *The Lancet*, 368(9533), pp. 387–403. Available at: [https://doi.org/10.1016/S0140-6736\(06\)69113-7](https://doi.org/10.1016/S0140-6736(06)69113-7).

Bonifati, V. (2014) ‘Genetics of Parkinson’s disease – state of the art, 2013’, *Parkinsonism & Related Disorders*, 20, pp. S23–S28. Available at: [https://doi.org/10.1016/S1353-8020\(13\)70009-9](https://doi.org/10.1016/S1353-8020(13)70009-9).

Bordoni, M. et al. (2022) ‘Lysosomes Dysfunction Causes Mitophagy Impairment in PBMCs of Sporadic ALS Patients’, *Cells*, 11(8), p. 1272. Available at: <https://doi.org/10.3390/cells11081272>.

Boylan, K. (2015) ‘Familial ALS’, *Neurologic clinics*, 33(4), pp. 807–830. Available at: <https://doi.org/10.1016/j.ncl.2015.07.001>.

Braun, S.M.G. and Jessberger, S. (2014) ‘Adult neurogenesis: mechanisms and functional significance’, *Development*, 141(10), pp. 1983–1986. Available at: <https://doi.org/10.1242/dev.104596>.

Breijyeh, Z. and Karaman, R. (2020) ‘Comprehensive Review on Alzheimer’s Disease: Causes and Treatment’, *Molecules*, 25(24), p. 5789. Available at: <https://doi.org/10.3390/molecules25245789>.

Cacace, R., Sleegers, K. and Van Broeckhoven, C. (2016) ‘Molecular genetics of early-onset Alzheimer’s disease revisited’, *Alzheimer’s & Dementia*, 12(6), pp. 733–748. Available at: <https://doi.org/10.1016/j.jalz.2016.01.012>.

Cang, C. et al. (2015) ‘TMEM175 Is an Organelle K(+) Channel Regulating Lysosomal Function’, *Cell*, 162(5), pp. 1101–1112. Available at: <https://doi.org/10.1016/j.cell.2015.08.002>.

Carlson, M. (2023) ‘org.Hs.eg.db: Genome wide annotation for Human’.

Charles A Janeway, J. et al. (2001) ‘The major histocompatibility complex and its functions’, in *Immunobiology: The Immune System in Health and Disease. 5th edition*. Garland Science. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK27156/> (Accessed: 19 August 2023).

Chaudhuri, K.R., Healy, D.G. and Schapira, A.H. (2006) ‘Non-motor symptoms of Parkinson’s disease: diagnosis and management’, *The Lancet Neurology*, 5(3), pp. 235–245. Available at: [https://doi.org/10.1016/S1474-4422\(06\)70373-8](https://doi.org/10.1016/S1474-4422(06)70373-8).

Chen, X., Guo, C. and Kong, J. (2012) ‘Oxidative stress in neurodegenerative diseases’, *Neural Regeneration Research*, 7(5), pp. 376–385. Available at: <https://doi.org/10.3969/j.issn.1673-5374.2012.05.009>.

Chen, Y.P. et al. (2013) ‘GAK rs1564282 and DGKQ rs11248060 increase the risk for Parkinson’s disease in a Chinese population’, *Journal of Clinical Neuroscience*, 20(6), pp. 880–883. Available at: <https://doi.org/10.1016/j.jocn.2012.07.011>.

Chen, Z. et al. (2021) ‘Revisiting the genome-wide significance threshold for common variant GWAS’, *G3 Genes|Genomes|Genetics*, 11(2), p. jkaa056. Available at: <https://doi.org/10.1093/g3journal/jkaa056>.

Chia, R. et al. (2021) ‘Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture’, *Nature Genetics*, 53(3), pp. 294–303. Available at: <https://doi.org/10.1038/s41588-021-00785-3>.

- Chial, H. (2008) ‘The Discovery of the Huntington Gene’, *Nature Education*, 1(1), p. 71.
- Chin, K.S., Teodorczuk, A. and Watson, R. (2019) ‘Dementia with Lewy bodies: Challenges in the diagnosis and management’, *Australian & New Zealand Journal of Psychiatry*, 53(4), pp. 291–303. Available at: <https://doi.org/doi.org/10.1177/0004867419835029>.
- Chitnis, T. and Weiner, H.L. (2017) ‘CNS inflammation and neurodegeneration’, *The Journal of Clinical Investigation*, 127(10), pp. 3577–3587. Available at: <https://doi.org/10.1172/JCI90609>.
- Cicardi, M.E. et al. (2021) ‘Proteostatic imbalance and protein spreading in amyotrophic lateral sclerosis’, *The EMBO Journal*, 40(10), p. e106389. Available at: <https://doi.org/10.15252/embj.2020106389>.
- Colletti, T. et al. (2021) ‘Prognostic Role of CSF β-amyloid 1–42/1–40 Ratio in Patients Affected by Amyotrophic Lateral Sclerosis’, *Brain Sciences*, 11(3), p. 302. Available at: <https://doi.org/10.3390/brainsci11030302>.
- Corder, E.H. et al. (1994) ‘Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease’, *Nature Genetics*, 7(2), pp. 180–184. Available at: <https://doi.org/10.1038/ng0694-180>.
- Corrochano, S. et al. (2012) ‘α-Synuclein levels modulate Huntington’s disease in mice’, *Human Molecular Genetics*, 21(3), pp. 485–494. Available at: <https://doi.org/10.1093/hmg/ddr477>.
- Crayton, H.J. and Rossman, H.S. (2006) ‘Managing the symptoms of multiple sclerosis: A multimodal approach’, *Clinical Therapeutics*, 28(4), pp. 445–460. Available at: <https://doi.org/10.1016/j.clinthera.2006.04.005>.
- Dehay, B. et al. (2010) ‘Pathogenic Lysosomal Depletion in Parkinson’s Disease’, *Journal of Neuroscience*, 30(37), pp. 12535–12544. Available at: <https://doi.org/10.1523/JNEUROSCI.1920-10.2010>.
- DeTure, M.A. and Dickson, D.W. (2019) ‘The neuropathological diagnosis of Alzheimer’s disease’, *Molecular Neurodegeneration*, 14(1), p. 32. Available at: <https://doi.org/10.1186/s13024-019-0333-5>.
- Dickson, D.W. et al. (2018) ‘APOE ε4 is associated with severity of Lewy body pathology independent of Alzheimer pathology’, *Neurology*, 91(12), pp. e1182–e1195. Available at: <https://doi.org/10.1212/WNL.0000000000006212>.
- Didonna, A. and Oksenberg, J.R. (2017) ‘The Genetics of Multiple Sclerosis’, in I.S. Zagon and P.J. McLaughlin (eds) *Multiple Sclerosis: Perspectives in Treatment and Pathogenesis*. Brisbane (AU): Codon Publications. Available at: <http://www.ncbi.nlm.nih.gov/books/NBK470155/> (Accessed: 25 May 2023).
- Do, C.B. et al. (2011) ‘Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson’s disease’, *PLoS genetics*, 7(6), p. e1002141. Available at: <https://doi.org/10.1371/journal.pgen.1002141>.

Faissner, S. *et al.* (2019) ‘Progressive multiple sclerosis: from pathophysiology to therapeutic strategies’, *Nature Reviews Drug Discovery*, 18(12), pp. 905–922. Available at: <https://doi.org/10.1038/s41573-019-0035-2>.

Fang, C. *et al.* (2020) ‘Cognition Deficits in Parkinson’s Disease: Mechanisms and Treatment’, *Parkinson’s Disease*, 2020, p. 2076942. Available at: <https://doi.org/10.1155/2020/2076942>.

Federico, A. *et al.* (2012) ‘Mitochondria, oxidative stress and neurodegeneration’, *Journal of the Neurological Sciences*, 322(1), pp. 254–262. Available at: <https://doi.org/10.1016/j.jns.2012.05.030>.

Foguem, C. and Manckoundia, P. (2018) ‘Lewy Body Disease: Clinical and Pathological “Overlap Syndrome” Between Synucleinopathies (Parkinson Disease) and Tauopathies (Alzheimer Disease)’, *Current Neurology and Neuroscience Reports*, 18(5), p. 24. Available at: <https://doi.org/10.1007/s11910-018-0835-5>.

Gallo, V. *et al.* (2019) ‘Exploring causality of the association between smoking and Parkinson’s disease’, *International Journal of Epidemiology*, 48(3), pp. 912–925. Available at: <https://doi.org/10.1093/ije/dyy230>.

Garcia-Esparcia, P. *et al.* (2017) ‘Dementia with Lewy Bodies: Molecular Pathology in the Frontal Cortex in Typical and Rapidly Progressive Forms’, *Frontiers in Neurology*, 8, p. 89. Available at: <https://doi.org/10.3389/fneur.2017.00089>.

Gatto, E.M. *et al.* (2020) ‘Huntington disease: Advances in the understanding of its mechanisms’, *Clinical Parkinsonism & Related Disorders*, 3, p. 100056. Available at: <https://doi.org/10.1016/j.prdoa.2020.100056>.

Ge, Y.-J. *et al.* (2023) ‘Prioritization of Drug Targets for Neurodegenerative Diseases by Integrating Genetic and Proteomic Data From Brain and Blood’, *Biological Psychiatry*, 93(9), pp. 770–779. Available at: <https://doi.org/10.1016/j.biopsych.2022.11.002>.

Ghasemi, N., Razavi, S. and Nikzad, E. (2017) ‘Multiple Sclerosis: Pathogenesis, Symptoms, Diagnoses and Cell-Based Therapy’, *Cell Journal (Yakhteh)*, 19(1), pp. 1–10.

Goldschmidt, H.L. *et al.* (2016) ‘DGK θ Catalytic Activity is Required for Efficient Recycling of Presynaptic Vesicles at Excitatory Synapses’, *Cell reports*, 14(2), pp. 200–207. Available at: <https://doi.org/10.1016/j.celrep.2015.12.022>.

Gómez-Pinedo, U. *et al.* (2016) ‘Immununochemical Markers of the Amyloid Cascade in the Hippocampus in Motor Neuron Diseases’, *Frontiers in Neurology*, 7. Available at: <https://www.frontiersin.org/articles/10.3389/fneur.2016.00195> (Accessed: 30 July 2023).

Gong, C.-X. and Iqbal, K. (2008) ‘Hyperphosphorylation of Microtubule-Associated Protein Tau: A Promising Therapeutic Target for Alzheimer Disease’, *Current Medicinal Chemistry*, 15(23), pp. 2321–2328.

Guella, I. *et al.* (2016) ‘ α -synuclein genetic variability: A biomarker for dementia in Parkinson disease’, *Annals of Neurology*, 79(6), pp. 991–999. Available at: <https://doi.org/10.1002/ana.24664>.

Guo, Y. *et al.* (2011) ‘HLA rs3129882 variant in Chinese Han patients with late-onset sporadic Parkinson disease’, *Neuroscience Letters*, 501(3), pp. 185–187. Available at: <https://doi.org/10.1016/j.neulet.2011.05.245>.

Gupta, A. and Kushwaha, S. (2019) ‘Maternally Inherited Huntington Disease and Paternally Inherited Neurofibromatosis Type-1 In an Individual- A Rare Co-Occurrence’, *Journal of Neurology & Neuromedicine*, 4(2). Available at: <https://www.jneurology.com/articles/maternally-inherited-huntington-disease-and-paternally-inherited-neurofibromatosis-type1-in-an-individual-a-rare-cooccurrence.html> (Accessed: 24 May 2023).

Haase, S. and Linker, R.A. (2021) ‘Inflammation in multiple sclerosis’, *Therapeutic Advances in Neurological Disorders*, 14, p. 17562864211007688. Available at: <https://doi.org/10.1177/17562864211007687>.

Haass, C. *et al.* (2012) ‘Trafficking and Proteolytic Processing of APP’, *Cold Spring Harbor Perspectives in Medicine*, 2(5), p. a006270. Available at: <https://doi.org/10.1101/cshperspect.a006270>.

Haass, C. and Strooper, B.D. (1999) ‘The Presenilins in Alzheimer’s Disease--Proteolysis Holds the Key’, *Science*, 286(5441), pp. 916–919.

Hamza, T.H. *et al.* (2010) ‘Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson’s disease’, *Nature Genetics*, 42(9), pp. 781–785. Available at: <https://doi.org/10.1038/ng.642>.

Hardy, J.A. and Higgins, G.A. (1992) ‘Alzheimer’s Disease: The Amyloid Cascade Hypothesis’, *Science*, 256(5054), pp. 184–185. Available at: <https://doi.org/10.1126/science.1566067>.

Hauser, S.L. *et al.* (1989) ‘Extended major histocompatibility complex haplotypes in patients with multiple sclerosis’, *Neurology*, 39(2 Pt 1), pp. 275–277. Available at: <https://doi.org/10.1212/wnl.39.2.275>.

Hayashi, Y., Homma, K. and Ichijo, H. (2016) ‘SOD1 in neurotoxicity and its controversial roles in SOD1 mutation-negative ALS’, *Advances in Biological Regulation*, 60, pp. 95–104. Available at: <https://doi.org/10.1016/j.jbior.2015.10.006>.

Hu, M. *et al.* (2022) ‘Parkinson’s disease-risk protein TMEM175 is a proton-activated proton channel in lysosomes’, *Cell*, 185(13), pp. 2292-2308.e20. Available at: <https://doi.org/10.1016/j.cell.2022.05.021>.

Huang, K.P. (1989) ‘The mechanism of protein kinase C activation’, *Trends in Neurosciences*, 12(11), pp. 425–432. Available at: [https://doi.org/10.1016/0166-2236\(89\)90091-x](https://doi.org/10.1016/0166-2236(89)90091-x).

IMSGC (2019) ‘Multiple Sclerosis Genomic Map implicates peripheral immune cells & microglia in susceptibility’, *Science (New York, N.Y.)*, 365(6460), p. eaav7188. Available at: <https://doi.org/10.1126/science.aav7188>.

International Multiple Sclerosis Genetics Consortium *et al.* (2011) ‘Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis’, *Nature*, 476(7359), pp. 214–219. Available at: <https://doi.org/10.1038/nature10251>.

James, L.M. and Georgopoulos, A.P. (2020) ‘Shared Human Leukocyte Antigen (HLA) Coverage in dementia and Parkinson’s disease’, *Journal of Neurology & Neuromedicine*, 5(3). Available at: <https://www.jneurology.com/articles/shared-human-leukocyte-antigen-hla-coverage-in-dementia-and-parkinsons-disease.html> (Accessed: 23 August 2023).

James, L.M. and Georgopoulos, A.P. (2021) ‘Immunogenetic Epidemiology of Dementia and Parkinson’s Disease in 14 Continental European Countries: Shared Human Leukocyte Antigen (HLA) Profiles’, *Journal of Immunological Sciences*, 5(2). Available at: <https://www.immunologyresearchjournal.com/articles/immunogenetic-epidemiology-of-dementia-and-parkinsons-disease-in-14-continentaleuropean-countries-shared-human-leukocyte-antigen-hla-profiles.html> (Accessed: 23 August 2023).

Jimenez-Sanchez, M. *et al.* (2017) ‘Huntington’s Disease: Mechanisms of Pathogenesis and Therapeutic Strategies’, *Cold Spring Harbor Perspectives in Medicine*, 7(7), p. a024240. Available at: <https://doi.org/10.1101/cshperspect.a024240>.

Kannarkat, G.T. *et al.* (2015) ‘Common genetic variant association with altered HLA expression, synergy with pyrethroid exposure, and risk for Parkinson’s disease: an observational and case–control study’, *npj Parkinson’s Disease*, 1(1), pp. 1–9. Available at: <https://doi.org/10.1038/npjparkd.2015.2>.

Kasper, L.H. and Shoemaker, J. (2010) ‘Multiple sclerosis immunology: The healthy immune system vs the MS immune system’, *Neurology*, 74 Suppl 1, pp. S2–8. Available at: <https://doi.org/10.1212/WNL.0b013e3181c97c8f>.

Kim, G.H. *et al.* (2015) ‘The Role of Oxidative Stress in Neurodegenerative Diseases’, *Experimental Neurobiology*, 24(4), pp. 325–340. Available at: <https://doi.org/10.5607/en.2015.24.4.325>.

Konno, S. *et al.* (2009) ‘Genetic Impact of a Butyrophilin-like 2 (BTNL2) Gene Variation on Specific IgE Responsiveness to Dermatophagoides farinae (Der f) in Japanese’, *Allergology International*, 58(1), pp. 29–35. Available at: <https://doi.org/10.2332/allergolint.08-OA-0005>.

Korn, T. (2008) ‘Pathophysiology of multiple sclerosis’, *Journal of Neurology*, 255(6), pp. 2–6. Available at: <https://doi.org/10.1007/s00415-008-6001-2>.

Kouchaki, E. *et al.* (2014) ‘Numerical status of CD4(+)CD25(+)FoxP3(+) and CD8(+)CD28(-) regulatory T cells in multiple sclerosis’, *Iranian Journal of Basic Medical Sciences*, 17(4), pp. 250–255.

LaFerla, F.M., Green, K.N. and Oddo, S. (2007) ‘Intracellular amyloid- β in Alzheimer’s disease’, *Nature Reviews Neuroscience*, 8(7), pp. 499–509. Available at: <https://doi.org/10.1038/nrn2168>.

Lambert, J.-C. *et al.* (2013) ‘Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease’, *Nature Genetics*, 45(12), pp. 1452–1458. Available at: <https://doi.org/10.1038/ng.2802>.

Latimer, C.S. and Montine, T.J. (2021) *Epidemiology, pathology, and pathogenesis of dementia with Lewy bodies - UpToDate*, UpToDate. Available at: <https://www.uptodate.com/contents/epidemiology-pathology-and-pathogenesis-of-dementia-with-lewy-bodies> (Accessed: 3 May 2023).

Lee, J.-H. *et al.* (2010) ‘Lysosomal Proteolysis and Autophagy Require Presenilin 1 and Are Disrupted by Alzheimer-Related PS1 Mutations’, *Cell*, 141(7), pp. 1146–1158. Available at: <https://doi.org/10.1016/j.cell.2010.05.008>.

Lenormand, C. *et al.* (2012) ‘HLA-DQA2 and HLA-DQB2 Genes Are Specifically Expressed in Human Langerhans Cells and Encode a New HLA Class II Molecule’, *The Journal of Immunology*, 188(8), pp. 3903–3911. Available at: <https://doi.org/10.4049/jimmunol.1103048>.

Lill, C.M. (2016) ‘Genetics of Parkinson’s disease’, *Molecular and Cellular Probes*, 30(6), pp. 386–396. Available at: <https://doi.org/10.1016/j.mcp.2016.11.001>.

Lim, E.W. *et al.* (2019) ‘Amyloid- β and Parkinson’s disease’, *Journal of Neurology*, 266(11), pp. 2605–2619. Available at: <https://doi.org/10.1007/s00415-018-9100-8>.

Lu, R.-C. *et al.* (2017) ‘Association of HLA-DRB1 polymorphism with Alzheimer’s disease: a replication and meta-analysis’, *Oncotarget*, 8(54), pp. 93219–93226. Available at: <https://doi.org/10.18632/oncotarget.21479>.

Lyketsos, C.G. *et al.* (2002) ‘Prevalence of Neuropsychiatric Symptoms in Dementia and Mild Cognitive ImpairmentResults From the Cardiovascular Health Study’, *JAMA*, 288(12), pp. 1475–1483. Available at: <https://doi.org/10.1001/jama.288.12.1475>.

Mackenzie, I.R.A. and Munoz, D.G. (1998) ‘Nonsteroidal anti-inflammatory drug use and Alzheimer-type pathology in aging’, *Neurology*, 50(4), pp. 986–990. Available at: <https://doi.org/10.1212/WNL.50.4.986>.

Madabhushi, R., Pan, L. and Tsai, L.-H. (2014) ‘DNA Damage and Its Links to Neurodegeneration’, *Neuron*, 83(2), pp. 266–282. Available at: <https://doi.org/10.1016/j.neuron.2014.06.034>.

Malm, T.M. *et al.* (2008) ‘Minocycline reduces engraftment and activation of bone marrow-derived cells but sustains their phagocytic activity in a mouse model of Alzheimer’s disease’, *Glia*, 56(16), pp. 1767–1779. Available at: <https://doi.org/10.1002/glia.20726>.

Mangalam, A.K., Taneja, V. and David, C.S. (2013) ‘HLA Class II Molecules Influence Susceptibility versus Protection in Inflammatory Diseases by Determining the Cytokine Profile’, *The Journal of Immunology*, 190(2), pp. 513–519. Available at: <https://doi.org/10.4049/jimmunol.1201891>.

Marras, C. *et al.* (2018) ‘Prevalence of Parkinson’s disease across North America’, *npj Parkinson’s Disease*, 4(1), pp. 1–7. Available at: <https://doi.org/10.1038/s41531-018-0058-0>.

Mathys, H. *et al.* (2019) ‘Single-cell transcriptomic analysis of Alzheimer’s disease’, *Nature*, 570(7761), pp. 332–337. Available at: <https://doi.org/10.1038/s41586-019-1195-2>.

Mazzulli, J.R. *et al.* (2011) ‘Gaucher Disease Glucocerebrosidase and α -Synuclein Form a Bidirectional Pathogenic Loop in Synucleinopathies’, *Cell*, 146(1), pp. 37–52. Available at: <https://doi.org/10.1016/j.cell.2011.06.001>.

McCord, J.M. and Fridovich, I. (1969) ‘Superoxide Dismutase: AN ENZYMIC FUNCTION FOR ERYTHROCUPREIN (HEMOCUPREIN)’, *Journal of Biological Chemistry*, 244(22), pp. 6049–6055. Available at: [https://doi.org/10.1016/S0021-9258\(18\)63504-5](https://doi.org/10.1016/S0021-9258(18)63504-5).

McGeer, P.L. and McGeer, E.G. (2002) ‘Inflammatory processes in amyotrophic lateral sclerosis’, *Muscle & Nerve*, 26(4), pp. 459–470. Available at: <https://doi.org/10.1002/mus.10191>.

McGeer, P.L., McGeer, E.G. and Yasoijima, K. (2000) ‘Alzheimer disease and neuroinflammation’, in K. Jellinger, R. Schmidt, and M. Windisch (eds) *Advances in Dementia Research*. Vienna: Springer, pp. 53–57. Available at: https://doi.org/10.1007/978-3-7091-6781-6_8.

McKinney, W. (2010) ‘Data Structures for Statistical Computing in Python’, in. *Python in Science Conference*, Austin, Texas, pp. 56–61. Available at: <https://doi.org/10.25080/Majora-92bf1922-00a>.

McLean, C.Y. *et al.* (2010) ‘GREAT improves functional interpretation of cis-regulatory regions’, *Nature Biotechnology*, 28(5), pp. 495–501. Available at: <https://doi.org/10.1038/nbt.1630>.

Mejzini, R. *et al.* (2019) ‘ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now?’, *Frontiers in Neuroscience*, 13, p. 1310. Available at: <https://doi.org/10.3389/fnins.2019.01310>.

Minty, A. *et al.* (1993) ‘Interleukin-13 is a new human lymphokine regulating inflammatory and immune responses’, *Nature*, 362(6417), pp. 248–250. Available at: <https://doi.org/10.1038/362248a0>.

Misra, M.K., Damotte, V. and Hollenbach, J.A. (2018) ‘The immunogenetics of neurological disease’, *Immunology*, 153(4), pp. 399–414. Available at: <https://doi.org/10.1111/imm.12869>.

Monsonego, A. *et al.* (2003) ‘Increased T cell reactivity to amyloid β protein in older humans and patients with Alzheimer disease’, *The Journal of Clinical Investigation*, 112(3), pp. 415–422. Available at: <https://doi.org/10.1172/JCI18104>.

Morimoto, N. *et al.* (2007) ‘Increased autophagy in transgenic mice with a G93A mutant SOD1 gene’, *Brain Research*, 1167, pp. 112–117. Available at: <https://doi.org/10.1016/j.brainres.2007.06.045>.

Moss, D.J.H. *et al.* (2017) ‘Identification of genetic variants associated with Huntington’s disease progression: a genome-wide association study’, *The Lancet Neurology*, 16(9), pp. 701–711. Available at: [https://doi.org/10.1016/S1474-4422\(17\)30161-8](https://doi.org/10.1016/S1474-4422(17)30161-8).

Moulton, P.V. and Yang, W. (2012) ‘Air Pollution, Oxidative Stress, and Alzheimer’s Disease’, *Journal of Environmental and Public Health*, 2012, p. 472751. Available at: <https://doi.org/10.1155/2012/472751>.

Naito, S. *et al.* (1972) ‘Multiple Sclerosis: Association with HL—A3’, *Tissue Antigens*, 2(1), pp. 1–4. Available at: <https://doi.org/10.1111/j.1399-0039.1972.tb00111.x>.

Nalls, M.A. *et al.* (2019) ‘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), pp. 1091–1102. Available at: [https://doi.org/10.1016/S1474-4422\(19\)30320-5](https://doi.org/10.1016/S1474-4422(19)30320-5).

Nance, M.A. (2017) ‘Chapter 1 - Genetics of Huntington disease’, in A.S. Feigin and K.E. Anderson (eds) *Handbook of Clinical Neurology*. Elsevier (Huntington Disease), pp. 3–14. Available at: <https://doi.org/10.1016/B978-0-12-801893-4.00001-8>.

Nguyen, T. *et al.* (2006) ‘BTNL2, a Butyrophilin-Like Molecule That Functions to Inhibit T Cell Activation’, *Journal of Immunology (Baltimore, Md. : 1950)*, 176(12), pp. 7354–7360.

Nichols, E. *et al.* (2022) ‘Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019’, *The Lancet Public Health*, 7(2), pp. e105–e125. Available at: [https://doi.org/10.1016/S2468-2667\(21\)00249-8](https://doi.org/10.1016/S2468-2667(21)00249-8).

Nixon, R.A. *et al.* (2005) ‘Extensive Involvement of Autophagy in Alzheimer Disease: An Immuno-Electron Microscopy Study’, *Journal of Neuropathology & Experimental Neurology*, 64(2), pp. 113–122. Available at: <https://doi.org/10.1093/jnen/64.2.113>.

Nixon, R.A. (2007) ‘Autophagy, amyloidogenesis and Alzheimer disease’, *Journal of Cell Science*, 120(23), pp. 4081–4091. Available at: <https://doi.org/10.1242/jcs.019265>.

Ohene, A. (2018) *Parkinson’s prevalence expected to increase by 18% in next seven years, Parkinson’s Life*. Available at: <https://parkinsonslife.eu/parkinsons-prevalence-expected-to-increase-by-18-by-2025/> (Accessed: 1 May 2023).

Oliveros, J.C. (2015) *Venny 2.1.0, An interactive tool for comparing lists with Venn’s diagrams*. Available at: <https://bioinfogp.cnb.csic.es/tools/venny/> (Accessed: 15 August 2023).

Orgogozo, J.-M. *et al.* (2003) ‘Subacute meningoencephalitis in a subset of patients with AD after A β 42 immunization’, *Neurology*, 61(1), pp. 46–54. Available at: <https://doi.org/10.1212/01.WNL.0000073623.84147.A8>.

Pankratz, N. *et al.* (2009) ‘Genomewide association study for susceptibility genes contributing to familial Parkinson disease’, *Human Genetics*, 124(6), pp. 593–605. Available at: <https://doi.org/10.1007/s00439-008-0582-9>.

Paolo, G.D. *et al.* (2004) ‘Impaired PtdIns(4,5)P₂ synthesis in nerve terminals produces defects in synaptic vesicle trafficking’, *Nature*, 431(7007), pp. 415–422. Available at: <https://doi.org/10.1038/nature02896>.

Pardillo-Díaz, R. *et al.* (2022) ‘Oxidative Stress as a Potential Mechanism Underlying Membrane Hyperexcitability in Neurodegenerative Diseases’, *Antioxidants*, 11(8), p. 1511. Available at: <https://doi.org/10.3390/antiox11081511>.

Patel, J. and Balabanov, R. (2012) ‘Molecular Mechanisms of Oligodendrocyte Injury in Multiple Sclerosis and Experimental Autoimmune Encephalomyelitis’, *International Journal of Molecular Sciences*, 13(8), pp. 10647–10659. Available at: <https://doi.org/10.3390/ijms130810647>.

Patsopoulos, N.A. (2018) ‘Genetics of Multiple Sclerosis: An Overview and New Directions’, *Cold Spring Harbor Perspectives in Medicine*, 8(7), p. a028951. Available at: <https://doi.org/10.1101/cshperspect.a028951>.

Perdigoto, C.N. (2022) ‘Parkinson’s disease risk protein TMEM175 keeps lysosomes running on a proton leak’, *Nature Structural & Molecular Biology*, 29(7), pp. 626–626. Available at: <https://doi.org/10.1038/s41594-022-00809-4>.

Raymond, L.A. *et al.* (2011) ‘Pathophysiology of Huntington’s disease: time-dependent alterations in synaptic and receptor function’, *Neuroscience*, 198, pp. 252–273. Available at: <https://doi.org/10.1016/j.neuroscience.2011.08.052>.

Rhee, J.S. *et al.* (2002) ‘Beta phorbol ester- and diacylglycerol-induced augmentation of transmitter release is mediated by Munc13s and not by PKCs’, *Cell*, 108(1), pp. 121–133. Available at: [https://doi.org/10.1016/s0092-8674\(01\)00635-3](https://doi.org/10.1016/s0092-8674(01)00635-3).

van Rheenen, W. *et al.* (2021) ‘Common and rare variant association analyses in amyotrophic lateral sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology’, *Nature Genetics*, 53(12), pp. 1636–1648. Available at: <https://doi.org/10.1038/s41588-021-00973-1>.

Rodrigo, L.M. and Nyholt, D.R. (2021) ‘Imputation and Reanalysis of ExomeChip Data Identifies Novel, Conditional and Joint Genetic Effects on Parkinson’s Disease Risk’, *Genes*, 12(5), p. 689. Available at: <https://doi.org/10.3390/genes12050689>.

Sait, A. *et al.* (2021) ‘Viral Involvement in Alzheimer’s Disease’, *ACS Chemical Neuroscience*, 12(7), pp. 1049–1060. Available at: <https://doi.org/10.1021/acschemneuro.0c00719>.

Sasaki, S. (2011) ‘Autophagy in Spinal Cord Motor Neurons in Sporadic Amyotrophic Lateral Sclerosis’, *Journal of Neuropathology & Experimental Neurology*, 70(5), pp. 349–359. Available at: <https://doi.org/10.1097/NEN.0b013e3182160690>.

Sayers, E.W. *et al.* (2022) ‘Database resources of the national center for biotechnology information’, *Nucleic Acids Research*, 50(D1), pp. D20–D26. Available at: <https://doi.org/10.1093/nar/gkab1112>.

Schmidt, H., Williamson, D. and Ashley-Koch, A. (2007) ‘HLA-DR15 haplotype and multiple sclerosis: a HuGE review’, *American Journal of Epidemiology*, 165(10), pp. 1097–1109. Available at: <https://doi.org/10.1093/aje/kwk118>.

Schoenborn, J. and Wilson, C. (2007) ‘Regulation of Interferon- γ During Innate and Adaptive Immune Responses - ScienceDirect’, *Advances in Immunology*, 96, pp. 41–101.

Schumacher, J. *et al.* (2021) ‘Dementia with Lewy bodies: association of Alzheimer pathology with functional connectivity networks’, *Brain*, 144(10), pp. 3212–3225. Available at: <https://doi.org/10.1093/brain/awab218>.

Shellikeri, S. *et al.* (2017) ‘The neuropathological signature of bulbar-onset ALS: A systematic review’, *Neuroscience & Biobehavioral Reviews*, 75, pp. 378–392. Available at: <https://doi.org/10.1016/j.neubiorev.2017.01.045>.

Simón-Sánchez, J. *et al.* (2011) ‘Genome-wide association study confirms extant PD risk loci among the Dutch’, *European Journal of Human Genetics*, 19(6), pp. 655–661. Available at: <https://doi.org/10.1038/ejhg.2010.254>.

Smirnov, D.S. *et al.* (2020) ‘Cognitive decline profiles differ in Parkinson disease dementia and dementia with Lewy bodies’, *Neurology*, 94(20), pp. e2076–e2087. Available at: <https://doi.org/10.1212/WNL.0000000000009434>.

Smith, L. and Schapira, A.H.V. (2022) ‘GBA Variants and Parkinson Disease: Mechanisms and Treatments’, *Cells*, 11(8), p. 1261. Available at: <https://doi.org/10.3390/cells11081261>.

Sollis, E. *et al.* (2023) ‘The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource’, *Nucleic Acids Research*, 51(D1), pp. D977–D985. Available at: <https://doi.org/10.1093/nar/gkac1010>.

Srinivasan, E. *et al.* (2021) ‘Alpha-Synuclein Aggregation in Parkinson’s Disease’, *Frontiers in Medicine*, 8. Available at: <https://www.frontiersin.org/articles/10.3389/fmed.2021.736978> (Accessed: 1 May 2023).

Steele, N.Z.R. *et al.* (2017) ‘Fine-mapping of the human leukocyte antigen locus as a risk factor for Alzheimer disease: A case–control study’, *PLOS Medicine*, 14(3), p. e1002272. Available at: <https://doi.org/10.1371/journal.pmed.1002272>.

Stelzer, G. *et al.* (2016) ‘The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses’, *Current Protocols in Bioinformatics*, 54(1), p. 1.30.1-1.30.33. Available at: <https://doi.org/10.1002/cpbi.5>.

Stillwell, W. (2016) ‘Chapter 5 - Membrane Polar Lipids’, in W. Stillwell (ed.) *An Introduction to Biological Membranes (Second Edition)*. Elsevier, pp. 63–87. Available at: <https://doi.org/10.1016/B978-0-444-63772-7.00005-1>.

Sullivan, P.M. *et al.* (2017) ‘Autophagy-Lysosome Dysfunction in Amyotrophic Lateral Sclerosis and Frontotemporal Lobar Degeneration’, in *Lysosomes - Associated Diseases and Methods to Study Their Function*. IntechOpen. Available at: <https://doi.org/10.5772/intechopen.69371>.

Sveinbjornsdottir, S. (2016) ‘The clinical symptoms of Parkinson’s disease’, *Journal of Neurochemistry*, 139(S1), pp. 318–324. Available at: <https://doi.org/10.1111/jnc.13691>.

Sweeney, P. *et al.* (2017) ‘Protein misfolding in neurodegenerative diseases: implications and strategies’, *Translational Neurodegeneration*, 6(1), p. 6. Available at: <https://doi.org/10.1186/s40035-017-0077-5>.

Tang, F. *et al.* (2021) ‘Role of Munc18-1 in the biological functions and pathogenesis of neurological disorders’, *Molecular Medicine Reports*, 23(3), p. 198. Available at: <https://doi.org/10.3892/mmr.2021.11837>.

Tanigawa, Y., Dyer, E.S. and Bejerano, G. (2022) ‘WhichTF is functionally important in your open chromatin data?’, *PLoS computational biology*, 18(8), p. e1010378. Available at: <https://doi.org/10.1371/journal.pcbi.1010378>.

Tanzi, R.E. (2012) ‘The Genetics of Alzheimer Disease’, *Cold Spring Harbor Perspectives in Medicine*, 2(10), p. a006296. Available at: <https://doi.org/10.1101/cshperspect.a006296>.

Taupin, V. *et al.* (1997) ‘Increased severity of experimental autoimmune encephalomyelitis, chronic macrophage/microglial reactivity, and demyelination in transgenic mice producing tumor necrosis factor- α in the central nervous system’, *European Journal of Immunology*, 27(4), pp. 905–913. Available at: <https://doi.org/10.1002/eji.1830270416>.

Tewari, D. *et al.* (2021) ‘Role of Nitric Oxide in Neurodegeneration: Function, Regulation, and Inhibition’, *Current Neuropharmacology*, 19(2), pp. 114–126. Available at: <https://doi.org/10.2174/1570159X18666200429001549>.

The UniProt Consortium (2023) ‘UniProt: the Universal Protein Knowledgebase in 2023’, *Nucleic Acids Research*, 51(D1), pp. D523–D531. Available at: <https://doi.org/10.1093/nar/gkac1052>.

Thu, D.C.V. *et al.* (2010) ‘Cell loss in the motor and cingulate cortex correlates with symptomatology in Huntington’s disease’, *Brain*, 133(4), pp. 1094–1110. Available at: <https://doi.org/10.1093/brain/awq047>.

Tu-Sekine, B. and Raben, D.M. (2011) ‘Regulation and roles of neuronal diacylglycerol kinases: a lipid perspective’, *Critical Reviews in Biochemistry and Molecular Biology*, 46(5), pp. 353–364. Available at: <https://doi.org/10.3109/10409238.2011.577761>.

Uffelmann, E. *et al.* (2021) ‘Genome-wide association studies’, *Nature Reviews Methods Primers*, 1(1), pp. 1–21. Available at: <https://doi.org/10.1038/s43586-021-00056-9>.

Verghese, P.B., Castellano, J.M. and Holtzman, D.M. (2011) ‘Apolipoprotein E in Alzheimer’s disease and other neurological disorders’, *The Lancet Neurology*, 10(3), pp. 241–252. Available at: [https://doi.org/10.1016/S1474-4422\(10\)70325-2](https://doi.org/10.1016/S1474-4422(10)70325-2).

Walton, C. *et al.* (2020) ‘Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition’, *Multiple Sclerosis (Hounds Mills, Basingstoke, England)*, 26(14), pp. 1816–1821. Available at: <https://doi.org/10.1177/1352458520970841>.

Wang, H. *et al.* (2011) ‘Smoking and Risk of Amyotrophic Lateral Sclerosis: A Pooled Analysis of 5 Prospective Cohorts’, *Archives of Neurology*, 68(2), pp. 207–213. Available at: <https://doi.org/10.1001/archneurol.2010.367>.

Wang, H. *et al.* (2021) ‘Genetic and environmental factors in Alzheimer’s and Parkinson’s diseases and promising therapeutic intervention via fecal microbiota transplantation’, *npj Parkinson’s Disease*, 7(1), pp. 1–10. Available at: <https://doi.org/10.1038/s41531-021-00213-7>.

Wang, J.Z. *et al.* (2007) ‘A new method to measure the semantic similarity of GO terms’, *Bioinformatics*, 23(10), pp. 1274–1281. Available at: <https://doi.org/10.1093/bioinformatics/btm087>.

Wang, Z.-X., Wan, Q. and Xing, A. (2020) ‘HLA in Alzheimer’s Disease: Genetic Association and Possible Pathogenic Roles’, *NeuroMolecular Medicine*, 22(4), pp. 464–473. Available at: <https://doi.org/10.1007/s12017-020-08612-4>.

WHO (2022) *Ageing and health*, World Health Organisation. Available at: <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health> (Accessed: 2 August 2023).

WHO (2023) *Dementia*. Available at: <https://www.who.int/news-room/fact-sheets/detail/dementia> (Accessed: 18 August 2023).

Wickham, H. and Grolemund, G. (2016) *R for Data Science*. 1st edn. O’Reilly Media.

Wightman, D.P. et al. (2021) ‘A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer’s disease’, *Nature Genetics*, 53(9), pp. 1276–1282. Available at: <https://doi.org/10.1038/s41588-021-00921-z>.

Wightman, D.P. et al. (2023) ‘The genetic overlap between Alzheimer’s disease, amyotrophic lateral sclerosis, Lewy body dementia, and Parkinson’s disease’, *Neurobiology of Aging*, 127, pp. 99–112. Available at: <https://doi.org/10.1016/j.neurobiolaging.2023.03.004>.

Wijesekera, L.C. and Nigel Leigh, P. (2009) ‘Amyotrophic lateral sclerosis’, *Orphanet Journal of Rare Diseases*, 4(1), p. 3. Available at: <https://doi.org/10.1186/1750-1172-4-3>.

Wirths, O. and Bayer, T.A. (2008) ‘Motor impairment in Alzheimer’s disease and transgenic Alzheimer’s disease mouse models’, *Genes, Brain and Behavior*, 7(s1), pp. 1–5. Available at: <https://doi.org/10.1111/j.1601-183X.2007.00373.x>.

Wu, T. et al. (2021) ‘clusterProfiler 4.0: A universal enrichment tool for interpreting omics data’, *The Innovation*, 2(3). Available at: <https://doi.org/10.1016/j.xinn.2021.100141>.

Xu, H. and Ren, D. (2015) ‘Lysosomal Physiology’, *Annual Review of Physiology*, 77(1), pp. 57–80. Available at: <https://doi.org/10.1146/annurev-physiol-021014-071649>.

Xu, L. et al. (2020) ‘Global variation in prevalence and incidence of amyotrophic lateral sclerosis: a systematic review and meta-analysis’, *Journal of Neurology*, 267(4), pp. 944–953. Available at: <https://doi.org/10.1007/s00415-019-09652-y>.

Yamashita, S. and Ando, Y. (2015) ‘Genotype-phenotype relationship in hereditary amyotrophic lateral sclerosis’, *Translational Neurodegeneration*, 4(1), p. 13. Available at: <https://doi.org/10.1186/s40035-015-0036-y>.

Yang, J.H. et al. (2022) ‘Therapeutic Advances in Multiple Sclerosis’, *Frontiers in Neurology*, 13. Available at: <https://www.frontiersin.org/articles/10.3389/fneur.2022.824926> (Accessed: 24 August 2023).

Yang, X. et al. (2017) ‘HLA-DRA/HLA-DRB5 polymorphism affects risk of sporadic ALS and survival in a southwest Chinese cohort’, *Journal of the Neurological Sciences*, 373, pp. 124–128. Available at: <https://doi.org/10.1016/j.jns.2016.12.055>.

Youssef, S.A. (2018) ‘Chapter 66 - Pathology of Brain Aging and Animal Models of Neurodegenerative Diseases’, in J.L. Ram and P.M. Conn (eds) *Conn’s Handbook of Models*

for Human Aging (Second Edition). Academic Press, pp. 899–908. Available at: <https://doi.org/10.1016/B978-0-12-811353-0.00066-X>.

Yu, G. (2020) ‘Gene Ontology Semantic Similarity Analysis Using GOSemSim’, in B.L. Kidder (ed.) *Stem Cell Transcriptional Networks: Methods and Protocols*. New York, NY: Springer US (Methods in Molecular Biology), pp. 207–215. Available at: https://doi.org/10.1007/978-1-0716-0301-7_11.

Yu, W.H. *et al.* (2005) ‘Macroautophagy—a novel β-amyloid peptide-generating pathway activated in Alzheimer’s disease’, *Journal of Cell Biology*, 171(1), pp. 87–98. Available at: <https://doi.org/10.1083/jcb.200505082>.

Zarranz, J.J. *et al.* (2004) ‘The new mutation, E46K, of α-synuclein causes parkinson and Lewy body dementia’, *Annals of Neurology*, 55(2), pp. 164–173. Available at: <https://doi.org/10.1002/ana.10795>.

Zhang, Y. *et al.* (2002) ‘Selective cytotoxicity of intracellular amyloid β peptide1–42 through p53 and Bax in cultured primary human neurons’, *Journal of Cell Biology*, 156(3), pp. 519–529. Available at: <https://doi.org/10.1083/jcb.200110119>.

Zhu, J. and Paul, W.E. (2008) ‘CD4 T cells: fates, functions, and faults’, *Blood*, 112(5), pp. 1557–1569. Available at: <https://doi.org/10.1182/blood-2008-05-078154>.

Zivadinov, R. *et al.* (2007) ‘HLA-DRB1*1501, -DQB1*0301, -DQB1*0302, -DQB1*0602, and -DQB1*0603 alleles are associated with more severe disease outcome on MRI in patients with multiple sclerosis’, *International Review of Neurobiology*, 79, pp. 521–535. Available at: [https://doi.org/10.1016/S0074-7742\(07\)79023-2](https://doi.org/10.1016/S0074-7742(07)79023-2).

Appendix

Appendix A: scripts for data analysis

Scripts used for data analysis in this study can be accessed at:

https://github.com/khadijapaderwala/thesis_analysis.git