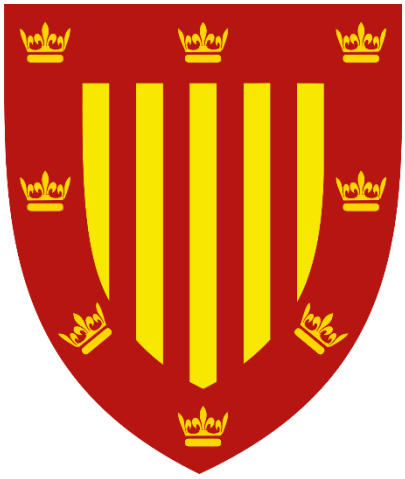
# Digitally quantified neuropathological correlates of structural and functional imaging biomarkers in progressive supranuclear palsy



**Tanrada Pansuwan**

Department of Clinical Neurosciences
University of Cambridge

This thesis is submitted for the degree of
*Doctor of Philosophy*

Peterhouse September 2023

# Declaration

I hereby declare that this thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text. It does not exceed the prescribed word limit (60,000 words) for the Clinical Medicine and Clinical Veterinary Medicine Degree Committee.

# Abstract

**Digitally quantified neuropathological correlates of structural and functional imaging biomarkers in progressive supranuclear palsy**

Tanrada Pansuwan

Neuroimaging measures are increasingly useful as *in vivo* biomarkers for differential diagnosis in neurodegenerative tauopathies. However, the relationship between imaging changes and neuropathology requires more thorough validation and beyond Alzheimer's disease. In tauopathies, tau is the key protein where specific pattern of tau aggregation and distribution can distinguish between different tauopathies. With this, progressive supranuclear palsy (PSP) is a prime disease for investigating the relationship between imaging changes and tau burden. Early studies have revealed that semi-quantitative pathological tau measures are predictive of *in vivo* atrophy but is less predictive of intrinsic functional connectivity. However, the semi-quantitative pathological measures used in these studies have limitations such that they are prone to inter-rater variability and may be insensitive to subtle pathological patterns. Therefore, in this thesis, I first attempted to develop an automated pipeline for quantifying total and tau positive cell density for neuronal and glial cells in *post mortem* samples. This promises to be a more objective, detailed, and scalable solution for pathological assessment. However, cell classification with high accuracy has proven to be very challenging; I then adapted the pipeline to only focus on quantifying tau type-specific aggregates in PSP. I used the pipeline to quantify PSP-related tau aggregates across cortical and subcortical regions and found a strong correspondence between the digitally quantified tau burden and the current consensus PSP staging scheme. I further assessed its clinicopathologic predictive power and found that total cortical tau and subcortical neurofibrillary densities correlate with clinical severity measured prior to death. Lastly, I assessed the relationship between both structural and functional imaging measures and *post mortem* tau type-specific density. I found that total tau density was associated with *in vivo* functional connectivity but not atrophy, suggesting that structural measure may be a less reliable marker of tau burden than functional measure in PSP. Overall, I have demonstrated that digitally quantified tau pathology can be a powerful tool to deepen our understanding of the role of tau for mechanistic studies and diagnosis in neurodegenerative tauopathies.

# Acknowledgement

I would firstly like to thank *Dr Timothy Rittman* for accepting me as his PhD student. I have learned so much throughout my 3 years of working with Tim and enjoyed every discussion we have had. I admire the dedication Tim has and the mission he is working towards despite all the challenges.

I would like to express my sincere gratitude towards *Jardine Foundation* and *Cambridge Trust* for sponsoring my PhD study, making my journey here at the University of Cambridge possible. Thank you, *Peterhouse*, for all the support throughout and the opportunity to meet numerous passionate and inspiring individuals.

Throughout my PhD, I have been working closely with *Dr Annelies Quaegebeur* in designing the digital pathology pipeline. Annelies has introduced me to the neuropathology world, which has been truly fascinating. I admire the work of neuropathologists, and the level of attention to detail required to work in this area. I would also like to thank *Professor James Rowe* for being very supportive, offering great advice and wider perspectives at every meeting. Also, thank you, *Dr Sanne Kaalund*, and *Dr Eric Hidari*, for making a start on the cell classification project together, you both have inspired me to tackle novel challenges with creativity, logical thinking, and optimism.

It has been amazing, having the chance to share scientific interests with my amazing lab members. Thank you, *Marion Peres*, for our work sessions and supporting each other throughout; *Dr David Whiteside*, for our conversations about statistics and functional connectivity; *Simon Jones*, for all the technical help with brain images. I am also grateful to my recent lab members, *Iryna, Marianna, Amir, Tanya*, and *Marcella* for their positive energy and insightful discussions.

My PhD journey would have been much harder without the support from loving friends. Many thanks to *New* and *Boom* for being my PhD buddies and helped me find my bearings in Cambridge. Unfortunately, I started my PhD during Covid-19 pandemic, special thanks to

*Mook, Arm, Oui, Joy and Sandy* for making Cambridge feel like home during those isolated times. Thank you, *Joe McCarthy,* and *Joe Allen*, for being amazing flatmates and friends. I would also like to thank my childhood friends for the much-needed support and love during tough times, especially *Pan*, *Mild* and *May*.

I am grateful to my family, especially my *mum* for her endless support both mentally and intellectually since my PhD application process. Thank you, *Onno*, for all the support, inspirations and tennis matches to keep me motivated. Thank you, *Manny*, for all the positive energy, enthusiasm and walks we have shared.

Thank you, all brain donors, participants, and all the relevant funding bodies for believing in our work. I hope this thesis has contributed some useful insights to the wider literature. Lastly, I would like to thank myself for pushing through until the end.

# Table of content

# List of figures

# List of tables

# List of abbreviations

1°M Primary motor 99

2D 2-dimensional 32

3D 3-dimensional 32

3R-tau 3-repeat-tau 21

4R-tau 4-repeated-tau 21

ACE-R Revised Addenbrooke's Cognitive Examination 28

AD Alzheimer's disease 22

AGD Argyrophilic Grain Disease 22

ARTAG Age-related tau astrogliopathy 140

Astro Astrocyte 62

BF Bayes' factor 51

BG Basal ganglia 99

bvFTD behavioural variant of frontotemporal dementia 20

CB coiled bodies 38

CBD Corticobasal Degeneration 22

Crl Credible interval 51

CTE Chronic Traumatic Encephalopathy 22

DAB 3,3'-diaminobenzidine 43

DN Dentate nucleus 99

FTD 53

FTLD-MAPT Frontotemporal lobar degeneration tau with MAPT mutation 21

GGT Globular glial tauopathy 22

GNR Glia:neuron ratio 56

PSPRS Progressive supranuclear palsy rating scale 28

PSP-RS PSP with Richardson syndrome 26

*PSP-SL* PSP with speech disorder 38

RBF Radial basis function 46

RBF-SVM Radial basis function SVM 46

RF Random forest 46

RFE Recursive feature elimination 44

ROPE Region of practical equivalence 51

STN Subthalamic nucleus 99

STR Striatum 99

SVM Support vector machine 46

TA tufted astrocytes 38

TF Tau fragments 97

# Chapter 1: Introduction

## 1.1. Overview

There have been advances in the fields of molecular pathology and neuroimaging in understanding neurodegenerative diseases. However, key questions remain about the relationship between molecular changes, brain networks and clinical syndromes. In this thesis, the focus is on improving our understanding of tau protein in humans using histology and imaging data. Therefore, in this chapter, neurodegenerative tauopathies as clinicopathological entities will firstly be reviewed to establish the significance of studying abnormal tau protein. Subsequently, important contributions from the field of molecular pathology in understanding tauopathies concerning the role of tau in normal health and disease, tauopathy classification schemes, and neuropathological examination will be explored. The motivation towards quantitative pathology and insights gained so far will then be evaluated. Next, the contribution from neuroimaging research on *in vivo* biomarker of tauopathies will be reviewed and the importance of establishing a bridge between molecular pathology and neuroimaging will be discussed. Progressive supranuclear palsy (PSP), a type of neurodegenerative diseases associated with tau neuropathology, is chosen as the focus of this thesis. A detailed explanation will be provided to justify this choice and to show that PSP has a potential to be a disease model for understanding tauopathies. Finally, the outline and contribution of this thesis towards the wider literature will be presented.

## 1.2. Neurodegenerative tauopathies

According to the World Health Organisation (World Health Organization, 2023), more than 55 million people worldwide have dementia, where 60-70% of these cases result from Alzheimer's disease. On a global scale, there are approximately 10 million new cases per annum and dementia is a major cause of disability and dependency in older people. In 2019, Alzheimer's disease and other forms of dementia were the 7$^{th}$ leading cause of global death and cost the global economy roughly 1.3 trillion US dollars. Dementia refers to a loss of cognition such as memory, language and problem-solving skills that ultimately affect one's

normal daily functions. There are several diseases that result in cognitive and motor neurodegenerative syndromes, one of which is a class of neurodegenerative diseases called ‘tauopathies’.

Tauopathies are characterised by abnormal tau protein accumulating as intracellular inclusions in neurons and/or glia, accompanied by neuronal loss and gliosis (Irwin, 2016; Kovacs, 2015). The clinical spectrum is broad and heterogenous, and often involves a dysexecutive cognitive impairment and parkinsonism, but may also cause language disorders, dystonia, eye movement abnormalities, and dysphagia (Höglinger et al., 2018). These symptoms reflect the brain regions affected and become more complex as more regions are affected to a greater degree (Forrest et al., 2019). Many clinical features can be seen in multiple tauopathies and other diseases, therefore they are not specific to a particular type of underlying pathology which makes in vivo diagnosis difficult (Höglinger et al., 2018; Kovacs, 2015). For example, 50% of the patients with clinical phenotype of behavioural variant of frontotemporal dementia (bvFTD) have tau-positive pathology of a wide range of underlying neuropathology including progressive supranuclear palsy, corticobasal degeneration, Pick’s disease and argyrophilic grain disease, or, they can have tau-negative frontotemporal degeneration, or Alzheimer’s disease (Höglinger et al., 2018; Irwin, 2016). Therefore, the traditional syndrome-based classification of tauopathies needs to be supplemented with details about molecular pathology to enable more specific classifications for better targeting of tau therapies (Höglinger et al., 2018).

## 1.3. Molecular pathology

Molecular pathology plays a key role in understanding neurodegenerative diseases as it permits the characterisation of neuropathological changes (Kovacs, 2019). This is crucial as such characterisation serves as a way to identify potential disease modifying targets (Guo & Lee, 2014; Iba et al., 2013). A molecular pathology approach has therefore been used to elucidate tau protein in health and disease, characterise tauopathies and examine neuropathology from *post mortem* samples for definite diagnosis and understanding disease progression.

### 1.3.1. What is tau?

Physiologically, tau can be found intracellularly and extracellularly where the function of extracellular tau is still unclear (Colin et al., 2020). Intracellular tau is predominantly found in the cytoplasm of neuronal axons and some can also be found in non-neuronal cells (Forrest et al., 2019). In neurons, the primary role of tau is to modulate the stability of axonal microtubules, by acting as a binding protein to support micro-tubule stabilization and axonal transport (Irwin, 2016). Extracellularly, tau can be found in the interstitial fluid and cerebrospinal fluid (Sexton et al., 2022).

Tau undergoes multiple post-translational modifications such as phosphorylation, acetylation, methylation, ubiquitination, and truncation (Sexton et al., 2022; Spillantini & Goedert, 2013; Wang & Mandelkow, 2016); abnormalities in these processes have been directly or indirectly implicated in neurodegeneration (Cohen et al., 2011; Irwin, 2016). Of the several modifications, phosphorylation is the major modification step that regulates the normal biological function of tau and has received the greatest interest in research (Irwin, 2016; Sexton et al., 2022; Y. Zhang et al., 2022). Tau hyperphosphorylation results in lower affinity between tau and axonal microtubules, thereby reducing the stability of the microtubule and affecting axonal transport and neurotransmission (Biernat et al., 2002; Sexton et al., 2022; Y. Zhang et al., 2022). The change in function resulting from tau hyperphosphorylation or other modifications is a result of conformational changes, resulting in insoluble misfolded tau aggregates which cause cellular dysfunction hence death (Sexton et al., 2022). In addition, the abnormal tau can be translocated from neuronal axons to the cell body and dendrites (Spillantini & Goedert, 2013). It has been proposed that abnormal tau may spread to other healthy cells, where it may promote further tau abnormality (Höglinger et al., 2018).

### 1.3.2. Characterising tauopathies

Tauopathies are largely sporadic but can result from mutations in the MAPT gene such as frontotemporal lobar degeneration tau with MAPT mutation (FTLD-MAPT) (Irwin, 2016).

Tau exists in six isoforms, produced by the alternative mRNA splicing on exons 2, 3, and 10 of the microtubule-associated protein tau (MAPT) gene, located on chromosome 17q21.3 (Andreadis, 2005). The isoforms include 3-repeat-tau (3R-tau) and 4-repeat-tau (4R-tau), with three isoforms in each group where the healthy human brain contains an equal ratio of

expressed 3R and 4R tau isoforms (Irwin, 2016). These are important, since different neurodegenerative diseases are associated with accumulation of distinct tau isoforms (Kovacs, 2015):

- **3R-tauopathies:** Pick's disease (PiD)
- **4R-tauopathies:** Progressive Supranuclear Palsy (PSP), Corticobasal Degeneration (CBD), Argyrophilic Grain Disease (AGD), Globular glial tauopathy (GGT)
- **3/4R tauopathies:** Alzheimer's disease (AD), Chronic Traumatic Encephalopathy (CTE).

A more recent study (Shi et al., 2021) using a cryo-electron microscopy to capture 3-dimensional structure of tau protein has found that tauopathies can be also characterized based on their disease-specific tau filament structures. This goes beyond the composition of 3R or 4R tau isoforms. Based on this, tauopathies can be classified based on 3 levels of classification. First, tauopathies are classified by the ordered tau protein core which also matches the classification by isoform composition (3R, 4R, 3/4R). At level 2, 4R tauopathies (but not others) can be separated into a 3-layered or 4-layered tau filament fold group, resulting in PSP and GGT in the former group, and CBD and AGD in the latter group. Lastly, CBD is differentiated from AGD, and PSP from GGT by the differences in the residue between the folds.

Table 1.1: Definitions and AT8 immunoreactive structures of neuronal, astroglial and oligodendroglial tau inclusion.

| Cell type | Immunoreactive structure | Definition |
|---|---|---|
| Neuron | - Pre-tangle / diffuse cytoplasmic inclusion | Diffuse fine granular staining of neuronal cytoplasm. |
| | - Neurofibrillary tangle | Fibrillar intracellular cytoplasm structures. |
| | - Pick's body | Cytoplasmic fibrillar spherical structures. |
| | - Threads | A segment of a thin neuronal process usually associated with axons. |
| | - Grains | 4–9 m spindle, coma or dot-like structures in the neuropil that are associated with dendrites. |
| | - Other spherical inclusions | Globular cytoplasmic inclusions that are various sized, and the staining pattern does not match the current definition of a Pick body. |
| Astrocyte | - Tufted | Star-like tufts of densely packed fibres in the proximal segments of astrocytic processes. |
| | - Astrocytic plaque | Annular cluster of short stubby lesions representing the distal segments of astrocytic processes. |
| | - Globular astroglial inclusion | Small globules in the astrocytic processes. |
| | - Ramified astrocyte | Tau positive astrocyte in PiD, with eccentric nuclei. |
| Oligodendrocyte | - Coiled body | Coil-like or coma-like intracytoplasmic profiles. |
| | - Globular oligodendroglial inclusion | Globular, spherical, or conical shaped structures with the diameter up to 15 μm. |

Note. Adapted from "Invited review: Neuropathology of tauopathies: principles and practice" by G. G. Kovacs, 2015, Journal of Neuropathology and Applied Neurobiology, p. 6. 

Table 1.2: Tau pathological hallmarks of neurodegenerative tauopathies in neuron, astrocyte, and oligodendrocyte under light microscope in specific vulnerable brain regions using AT8 stain.

| Tauopathy | Neuronal inclusion | Astrocytic inclusion | Oligodendroglia inclusion | Main anatomic vulnerability |
|---|---|---|---|---|
| *3R* | | | | |
| PiD | - Pick bodies<br>- Pick cells<br>- Neuropil threads | - Ramified astrocytes | - | - Frontal & temporal cortices<br>- Dentate gyrus<br>- Hippocampus |
| *4R* | | | | |
| PSP | - Globose-type neurofibrillary tangles<br>- Diffuse inclusion<br>- Neuropil threads | - Tufted astrocyte | - Coiled bodies | - Basal ganglia (specifically subthalamic nucleus, substantia nigra)<br>- Brainstem (specifically midbrain)<br>Frontal lobe |
| CBD | - Diffuse inclusion<br>- Ballooned neurons<br>- Neuropil threads<br>- Spherical inclusion | - Astrocytic plaque | - Coiled bodies | - Frontoparietal association cortices<br>- Neostriatum<br>Substantia nigra |
| AGD | - Spindle-shaped grains<br>- Diffuse inclusion<br>- Ballooned neurons | - Ramified astrocytes | - Coiled bodies | - Limbic structures |
| GGT | - Spherical inclusion<br>- Diffuse inclusion | - Globular astroglial inclusion | - Globular oligodendroglial inclusion | - Frontal, temporal lobes |
| *3/4R* | | | | |
| AD | - Flame-shaped neurofibrillary tangles<br>- Diffuse inclusion<br>- Neuritic plaques<br>- Neuropil threads | - | - | - Basal forebrain<br>- Frontal, temporal lobes<br>- Limbic structures<br>- Locus coeruleus<br>Olfactory bulb |
| CTE | - Neurofibrillary tangles | - Astrocytic tangles | - | - Frontal, temporal, and parietal lobes<br>- Depth of sulci and surrounding vasculature |
| PART | - Neurofibrillary tangles | - Diffuse inclusion | - Coiled bodies | - Medial temporal lobe<br>- Substantia nigra<br>- Brainstem |

Note. Adapted from "Pathology of Neurodegenerative diseases" by Brittany N Dugger and Dennis W. Dickson, 2017, Cold Spring Harb Perspect Biol, p. 2. Copyright 2017 Cold Spring Harbor Laboratory Press.

Numerous antibodies are available for staining hyperphosphorylated tau and identifying conformational modifications, with the commonest antibody for detecting hyperphosphorylated tau being AT8 that binds to the Ser202/Thr205 residue (Kovacs, 2015).

After staining, light microscopy can be used to inspect tau isoforms and immunoreactive structures for specific tauopathies where specific tauopathy can be classified based on the specific pattern of tau aggregation in the affected cell type and brain region (Table 1.1, Table 1.2).

Furthermore, tauopathies can also be described as primary or secondary tauopathies, depending on whether tau is the prominent feature of the pathology or not (Y. Zhang et al., 2022). For example; PiD, PSP, CBD, AGD are classified as primary tauopathies, whereas AD is a secondary tauopathy as amyloid beta is also involved and likely triggered the abnormal tau production (Chung et al., 2021; Irwin, 2016; Sexton et al., 2022).

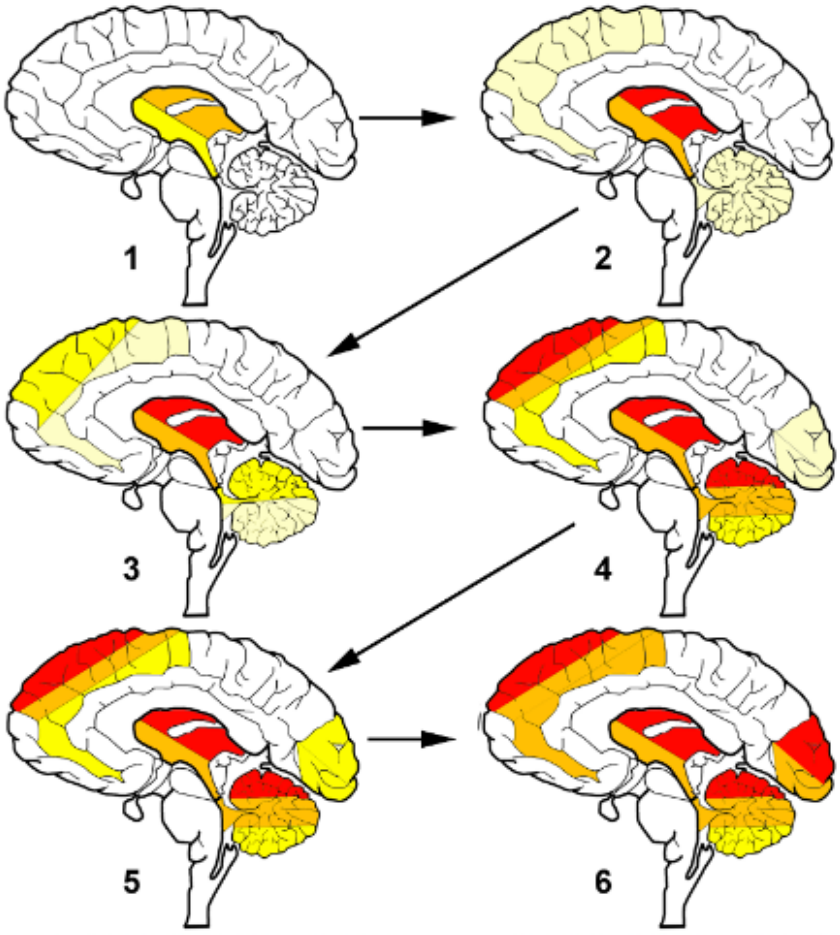
### 1.3.3. Neuropathological examination

*Post mortem* neuropathological examination is the gold standard approach for definite diagnosis in neurodegenerative diseases (Jack et al., 2018; Scheltens & Rockwood, 2011). This process involves the identification of disease-specific protein accumulation pattern in the affected cell type and brain region that correlates with specific diseases (see Table 1.2) (Dugger & Dickson, 2017; K. H. Kim et al., 2020; Scheltens & Rockwood, 2011). For example, amongst the 4R tauopathies, the morphology of astrocytic tau lesions is the key distinctive feature in differentiating between PSP and CBD such that tufted astrocytes are found only in PSP and astrocytic plaques are found only in CBD (Koga et al., 2021). Moreover, even though both PiD and AGD have common site of pathology at limbic structures, PiD shows wider spread of pathology to cortical structures, unlike AGD (Dugger & Dickson, 2017).

Identifying characteristic patterns of neuropathological progression has led to staging systems of neurodegenerative diseases to encapsulate disease progression (Colin et al., 2020). For example, Braak staging of AD (Braak et al., 2011; Braak & Braak, 1991, 1995) proposes that neurofibrillary changes first occur in the transentorhinal region in early stages of the disease before spreading to limbic areas and then to primary and secondary neocortical regions at the final stage. These staging systems enable a more detailed characterisation of individuals, and a framework within which to consider the sequential spreading pattern of neurodegenerative pathology (Rösler et al., 2019).

A recent PSP staging scheme has been proposed (Figure 1.1), based on tau type-specific burden (Kovacs et al., 2020). The PSP Richardson syndrome (PSP-RS) contains 6 stages, where each stage can be characterised by the severity of tau burden in the specific cell type and brain region. For example, tau burden accumulates first in the basal ganglia nuclei and dentate nucleus before spreading to cortical regions. The main cell type affected is different amongst the regions, such that the focus is on neuronal tau in the subthalamic nucleus, globus pallidus and dentate nucleus, while astrocytic tau is the focus in striatum. In the cortex, astroglial tau pathology is the focus, and the frontal lobe is affected before the occipital lobe. This staging scheme can help address questions about disease progression, including in different PSP subtypes.

Figure 1.1: The current PSP-RS staging scheme for neuropathological assessment.




| REGION | GP | | STN | | STR | | FR | | DE/CB | | OC | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Cell / Stage | N/O | | N | | A | | A | | N/O | | A | |
| 1 | +/++ | AND/OR | +/++ | AND/OR | +/++ | AND | - | AND | - | AND | - | |
| 2 | ++/+++ | AND | ++/+++ | AND | ++/+++ | AND | -/+ | AND/OR | -/+ | AND | - | |
| 3 | ++/+++ | AND | ++/+++ | AND | ++/+++ | AND | -/+<br>+<br>+ | AND | +<br>+<br>-/+ | AND | - | |
| 4 | ++/+++ | AND | ++/+++ | AND | ++/+++ | AND | +<br>++/+++<br>++/+++ | AND | ++/+++<br>++/+++<br>+ | AND | - | -/+ |
| 5 | ++/+++ | AND | ++/+++ | AND | ++/+++ | AND | +<br>++/+++<br>++/+++ | AND | ++/+++<br>++/+++<br>+ | AND | + | |
| 6 | ++/+++ | AND | ++/+++ | AND | ++/+++ | AND | ++/+++ | AND | ++/+++<br>++/+++<br>+ | AND | ++/+++ | |

Note. Kovacs et al (2020) noted that "−/ + Indicates single cell involvement; + indicates mild; + + / + + + indicates moderate/severe involvement. GP globus pallidus, STN subthalamic nucleus, STR striatum, FR frontal, DE/CB dentate nucleus and cerebellar white matter, OC occipital. This can be applied to all clinical subtypes. The evaluator should focus on different cell types in different brain regions: in GP and DE/CB neuronal (N) or oligodendroglial (O);

in the STN neuronal; in the STR and FR and OC cortices astroglial (A). The brain schema is a conceptual summary of the tabularized schema in the lower panel; thus the color coding of different brain regions reflect the variability in scores (or-or) required for a stage." (p. 115) This scheme was produced by Kovacs et al in 2020 for neuropathological testing criteria for PSP-RS. From "Distribution patterns of tau pathology in progressive supranuclear palsy", by Kovacs et al, 2020, *Acta neuropathologica*, *140*, 99-119. Copyright 2020 by The Author(s).

The current PSP-RS staging scheme has recently been validated in an independent cohort and is found to correlate with PSP rating scale (PSPRS) and revised Addenbrooke's Cognitive Examination (ACE-R) scores at death, supporting the association between pathological tau and cognitive impairment (Briggs et al., 2021). Nevertheless, there are limitations to the current staging scheme. Neuropathological examination is a semi-quantitative approach where visual inspection is performed to grade the severity of pathology on an ordinal scale (Kovacs, 2015; Kovacs et al., 2020; Scheltens & Rockwood, 2011). Defining a precise cut-off point for each stage is an artificial construct and can be challenging, as the underlying pathobiological processes of neurodegenerative disorders are dynamic (Del Tredici & Braak, 2020). Such examination is also inherently subjective as there are innate differences in visual perception and decision-making processes between individuals (Bera et al., 2019) and tissue preparation techniques can be different across institutions (Del Tredici & Braak, 2020; Jellinger, 2010; Litvan et al., 1996). These result in issues with inter- and intra-observer reliability (Alafuzoff et al., 2009; Scheltens & Rockwood, 2011). Pathologists also have to be extensively trained (Koga et al., 2021; Signaevsky et al., 2019), making the quality of assessment dependent on their level of experience. Lastly, given the laborious and time-consuming manual assessment process (Neltner et al., 2012; Walker et al., 2017), the current staging scheme is limited to a selection of regions which does not consider all the brain regions affected by PSP pathology. Advantages in digital pathology could potentially address a number of these issues.

## 1.4. Digital pathology

### 1.4.1. Towards digital pathology

Digital pathology promises to be a more objective method of assessing histology using computational approaches (Al-Janabi et al., 2012; Bera et al., 2019). The current semi-quantitative approach provides simple measurements suitable for routine autopsy. However, as larger amount of data with greater detail is required for research, digital pathology has a potential to address these current limitations. Such concepts have been around since 1960s (Prewitt & Mendelsohn, 1966) but has gained greater popularity due to the advancement in whole slide imaging at high resolution, and advances in artificial intelligence methods (Al-Janabi et al., 2012; Hamilton et al., 2014; Kayser et al., 2009).

Key advantages of digital pathology over the current standard practice include reducing inter-rater variability and increasing efficiency through automated pipelines which can be scaled to analyse larger datasets (Hamilton et al., 2014). This is crucial as more brain regions and data from multiple institutions can be analysed to ensure the generalisability of results (Tang et al., 2019) and a more comprehensive neuropathological characterisation of tauopathies would be possible (Khan et al., 2021). Moreover, as biological samples are complex, some pathological changes can be subtle and masked by the discrete nature of a semi-quantitative method (Walker et al., 2017). Therefore, novel pattern can potentially be captured by digital pathology to provide insights for mechanistic studies (Vega et al., 2021) and assist differential diagnosis by detecting disease-specific tau pathological hallmarks (Koga et al., 2021, 2022). In addition, as the standard semi-quantitative approach summaries pathology as a score, this limits the statistical power for comparative studies (Coughlin et al., 2022) and meaningful correlation with other measures of disease (Tang et al., 2019).

As richer information can be extracted using digital pathology, the main benefit will be towards enabling more research questions and complex analyses to be possible for mechanistic research. Digital pathology can also assist pathologists with autopsy verification by reducing time and expertise required from the hospitals. It is not designed to replace pathologists as human expertise will continue to be essential due to unforeseen decision-making and ethical reasons. Nevertheless, there are caveats and many challenges concerning

machine learning algorithm validation, interpretability, and standardisation (Bera et al., 2019; Hamilton et al., 2014; Vizcarra et al., 2020). If these limitations can be addressed, digital pathology has potential to address the need for more sensitive measures of disease burden.

### 1.4.2. What has digital pathology told us?

Within the last 5 years, studies using automated approaches to investigate neurodegenerative diseases can be grouped into 3 main categories: characterisation of tau aggregates in tauopathies, tau hallmark classification, and the practicality of automated pipeline for clinical use. In characterising tau aggregates, studies have strengthened our understanding of the relationship between tau burden and clinical symptoms (Kaalund et al., 2020; Marx et al., 2022). For example, Kaalund et al (2020) quantified neuronal loss and the proportion of neurons with hyperphosphorylated tau in the locus coeruleus of PSP patients. They used 'positive pixel count' method to count 'tau positive' cells with pixel values above the pre-determined tau positive threshold. With this, they found that PSP patients have a 49% reduction in noradrenergic neurons that contain neuromelanin relative to controls, and that 44% of noradrenergic neurons have hyperphosphorylated tau. Neuronal loss also negatively correlates with tau positive neurons in a non-linear fashion which is not due to age or disease duration and only neuronal loss but not tau burden is associated with clinical severity. These findings demonstrate the complex relationship between tau burden, neuronal loss, and clinical severity and that the integrity of locus coeruleus may mediate clinical severity independent of aging and prolonged disease duration effects, highlighting its potential for targeted pharmacological treatment.

Studies have also used automated pipelines to investigate tau aggregates in both cortical grey and white matter across tauopathies (Coughlin et al., 2022; Vega et al., 2021). This is powerful as white matter has substantial tau pathology but has not been well studied (Coughlin et al., 2022). Tau aggregates in white matter can be complex as they lack canonical phenotypes which can pose difficulties for visual inspection, but such a task is suitable for machine learning algorithms designed to analyse intricate patterns (Vega et al., 2021). With this, studies using automated approaches have found that grey and white matter tau correlated with each other in a disease-specific fashion (Coughlin et al., 2022; Vega et al., 2021). PSP and CBD showed equal distribution of tau in grey and white matter while tau predominantly

accumulates in grey matter in AD (Vega et al., 2021) which also correlates with cognitive score (Coughlin et al., 2022). Novel patterns of white matter tau aggregates were found which were morphologically distinct across AD, PSP, and CBD which could be useful for disease classification (Vega et al., 2021). Altogether, these studies have emphasised the importance of studying both grey and white matter for a better understanding of tau pathology in tauopathies and highlight the utility of digital pathology for repetitive and intricate tasks.

Due to a greater need for better stratification, characterization, and quantification of tau burden in tauopathies, studies have advanced from investigating tau burden as a whole to tau type-specific burden using a semi-quantitative approach (Jellinger, 2018; Kovacs et al., 2020). In developing further to reduce human error, and to increase accuracy and efficiency, there have been attempts to create automated tau type-specific quantification pipelines to aid differential pathological diagnosis (Koga et al., 2021, 2022; Signaevsky et al., 2019) and disease mechanism research (Marx et al., 2022; Tang et al., 2019). Studies focusing on classification tasks have shown high performance and have largely used convolution neural network (Signaevsky et al., 2019; Tang et al., 2019) and random forest approaches (Koga et al., 2022).

For differential diagnosis, studies have attempted to differentiate between disease-specific tau hallmarks (Koga et al., 2021, 2022). Koga et al (2021) created a pipeline to classify tufted astrocytes which are characteristic of PSP tau pathology, astrocytic plaques characteristic of CBD and neuritic plaques characteristic of AD. Such a pipeline would be useful for aiding neuropathologists in their decision-making for differential diagnosis between CBD and PSP, as well as informing about co-pathology as inferred from neuritic plaques. Similarly, Koga et al (2022) created a pipeline to classify 5 different types of tau burden to differentiate between 4 tauopathies which include PSP, CBD, PiD and AD with classification performance over 95%. This further demonstrates that automated pipelines can achieve high accuracy to support diagnosis in the clinical setting, applied to tauopathies. Furthermore, these digital pipelines (Koga et al., 2022; Marx et al., 2022; Signaevsky et al., 2019) output detailed objective quantification, which can be used for correlation studies such as correlating tau burden with clinicopathological, molecular and genetic indices to investigate novel research questions about their relationships (Signaevsky et al., 2019).

In mechanistic research using automated pipelines, neurofibrillary tangle (NFT) has received the greatest attention amongst all tau types (Marx et al., 2022; Signaevsky et al., 2019). This is likely because NFT is prominent across tauopathies, especially in AD which is the most common type of dementia (Y. Zhang et al., 2022). Braak staging of AD involves the assessment of NFT distribution and is commonly used for measuring pathological tau in research and clinical settings (Braak & Braak, 1995). However, Braak staging is not well validated in beta-amyloid negative disease and may lack clinicopathologic predictive power in those disorders (Iida et al., 2021). Marx et al (2022) used an AI-derived measure of NFT burden and Braak stage to study the relationship between NFT burden, aging, and clinical presentation in patients with PART. These patients are elderly individuals that have Alzheimer-type NFT but are amyloid-beta negative and may or may not have cognitive impairment. It remains unclear whether cognitive impairment in PART is a result of age-related neurodegenerative processes or not. They found that NFT counts but not Braak stage correlated with cognitive impairment in PART after controlling for age. This suggests that NFT burden which is not due to aging correlates with cognitive impairment in PART and highlights the advantage of using artificial intelligence derived metrics over a simplistic semi-quantitative stage-based approach for measuring NFT burden.

Taking it further, Yushkevich et al (2021) developed a method for quantitative 3D mapping of NFTs. They found a significant variation of NFT along the anterior-posterior axis of the brain that could be missed by the 2D approach as histological sectioning is usually done in a single (usually coronal) plane. Such quantification enables a more detailed characterisation of tau distribution in normal ageing and pathology and could also be useful for validating neuroimaging biomarkers (Ravikumar et al., 2021).

Lastly, with numerous studies developing automated pipelines, it is also important to consider practical factors that will influence adoption rate by researchers and healthcare professionals such as pipeline robustness and ease of use. Studies have begun to validate their pipeline on data from multiple institutions (Coughlin et al., 2022; Vizcarra et al., 2020; Yushkevich et al., 2021). This is important as tissue preparation standards can vary across institutions such as criteria for cohort selection, slide preparation techniques, *post mortem* interval, skills of the scientists (Alafuzoff et al., 2006; Vizcarra et al., 2020) and imaging systems (Bautista et al., 2014). Moreover, increasing efforts have been dedicated towards developing an easy-to-use interface that requires minimal computational skills from end users

(Jungo & Hewer, 2023; Tang et al., 2019). For example, Tang et al (2019) created a custom web interface for amyloid plaque annotation for pathologists to annotate objects at a faster rate using mouse and keystrokes where data is then stored in a standardised query language format. Jungo and Hewer (2023) assessed the utility of publicly available code-free machine learning platforms and found that Microsoft Custom Vision and Google AutoML could easily be used by pathologists in a code-free manner and still achieve high classification performance. These efforts will contribute towards ensuring the translation of a highly robust and accurate automated pipelines from development to real-world adoption by both researchers for mechanistic research and healthcare professionals to aid their decision-making processes in hospitals.

All in all, digital pathology is a promising discipline in improving our understanding of the pathogenesis of neurodegenerative tauopathies and aiding decision-making processes of pathologists and clinicians.

## 1.5. Neuroimaging

### 1.5.1. Neuroimaging as a biomarker

Neuroimaging plays an increasing role in identifying reliable and sensitive *in vivo* biomarkers for studies of neurodegenerative disease progression (Vogel et al., 2021), prognosis (Jo et al., 2019), mechanisms (Iturria-Medina et al., 2016) and disease-modifying treatments (Schott et al., 2010; Staffaroni et al., 2017).

Biomarkers are measurable characteristics of specific biological processes, indicative of normal or pathological processes (Strimbu & Tavel, 2010). They are important as clinical symptoms are not sensitive nor specific enough to characterise neurodegenerative diseases (Jack et al., 2018), because neuropathological changes can be present with few or no symptoms or with non-specific symptoms that overlap between diseases (Bennett et al., 2006; Knopman et al., 2003). Biomarkers are largely grouped into two categories; disease-specific biomarkers are useful for diagnosis and prognosis while biomarkers that reflect non-disease-specific processes provide additional insights into the underlying neuropathological changes (Jack et al., 2018). For example, beta-amyloid plaque and neurofibrillary tangle biomarkers together are specific to AD and are therefore used to determine if a person belongs to the AD

spectrum while neuronal injury biomarkers such as elevated total tau burden or decreased glucose metabolism are used to determine disease severity ("2018 Alzheimer's Disease Facts and Figures," 2018; Jack et al., 2018) as these changes also occurs in non-AD conditions (Kovacs et al., 2013). Imaging biomarkers that reflect the underlying neuropathological changes are of interest to define the disease (Rittman, 2020).

### 1.5.2. MRI as *in vivo* biomarker

Magnetic resonance imaging (MRI) is a non-invasive technique widely used in research and clinical practice to corroborate a diagnosis together with the presence of clinical syndromes and tracking disease progression (Rittman, 2020). In AD, MRI-based measures of atrophy have been shown to relate to upstream neuronal deficits (Whitwell et al., 2008) and downstream clinical symptoms (Grundman et al., 2002; Vemuri et al., 2009). The degree of medial temporal lobe atrophy is widely used to predict the progression to AD from mild cognitive impairment (MCI), presumed to reflect underlying neuronal loss (Frisoni et al., 2010). Medial temporal lobe atrophy is predictive of future AD in both non-demented patients and those with MCI; and this prediction improves when age and memory scores are added to medial temporal lobe atrophy measures (Visser et al., 2002). Various longitudinal studies have also shown that MCI patients that later develop AD have greater volumetric reduction in the hippocampal area (Chételat et al., 2005; Tapiola et al., 2008). Visual assessment of MRI scans using either rating scales (Harper et al., 2016) or automated approaches can also identify AD and non-AD tauopathies (Vemuri et al., 2011). Moreover, MRI-based measures of atrophy are included as mandatory and supportive features for the diagnosis of non-AD disorders. For example; extensive white matter changes are mandatory for the diagnosis of vascular dementia (Román et al., 1993) and focal frontal or temporal atrophy is supportive of frontotemporal dementia (Neary et al., 1998).

Functional MRI estimates changes in blood flow, reflective of activation in different brain regions during rest or when actively performing a task, which can be used to study functional brain networks (Bullmore & Sporns, 2009). Studies have shown that functional networks can be affected in a disease-specific fashion, for example, the default mode network which relates to episodic memory is affected in AD (Ingala et al., 2021; Seeley et al., 2009) and the emotional salience processing network is affected in bvFTD (Pasquini et al., 2020; Seeley et al., 2009). In general, even though functional measures show focal changes in dementia, they

are currently not being used in clinical practice as there are significant overlaps amongst diagnostic groups (Chouliaras & O'Brien, 2023).

Overall, structural MRI is the most widely used neuroimaging biomarker in clinical practice and integrated in various diagnostic and research guidelines for dementia as it is a practical and inexpensive measure that can improve diagnostic accuracy (Chouliaras & O'Brien, 2023; Harper et al., 2016). Nevertheless, MRI measures still have limitations as biomarkers because imaging measures have mostly been presumed to reflect the underlying pathology, with little autopsy verification (Jack et al., 2002).

## 1.6. Neuroimaging-pathology

### 1.6.1. Validating *in vivo* structural and functional biomarkers

Studies have first examined the underlying relationship between pathology and structural MRI biomarker by studying the relationship between brain volume and simple pathological measures. In AD, brain volume was correlated with Braak stages, counts or percentage area covered by neurons and pathological proteins (Apostolova et al., 2015; Burton et al., 2009; Jack et al., 2002; Whitwell et al., 2012). Atrophy in the hippocampus was found to correlate with the Braak stage (Jack et al., 2002) as well as tau, beta-amyloid burden and neuronal count (Apostolova et al., 2015). MTA correlates with the distribution of NFT in AD, more strongly than with amyloid plaques, or Lewy bodies in dementia with Lewy bodies (Burton et al., 2009). The pattern of atrophy across the cortex and medial temporal lobe can be used to distinguish between AD pathological subtypes that were grouped based on the distribution of NFT, suggesting that MRI measure can reliably track NFT burden in AD (Whitwell et al., 2012).

More recently, studies have investigated a range of neurological conditions, including functional biomarkers and a wider range of pathological measures such as neuronal loss and size, astrogliosis, microvacuolation, combined neuronal and glial tau inclusions (Jardim et al., 2016; E.-J. Kim, 2020; Popescu et al., 2015; Spina et al., 2019). These studies are essential in improving the specificity of assessing *in vivo* biomarkers and can also reveal patho-mechanisms. Grey matter volume has been shown to correlate with neuronal density and size

in multiple sclerosis (Popescu et al., 2015) and hippocampal atrophy also correlates with neuronal loss in epilepsy with hippocampal sclerosis, useful for surgical prognosis (Jardim et al., 2016). A study also (Spina et al., 2019) found that brain atrophy reflected an independent contribution from neurodegeneration and tau burden in PSP and CBD at the whole brain level. Specifically, atrophy in PSP was more driven by neuronal tau, whereas atrophy in CBD was more driven by glial tau pathology. Moreover, functional networks showed less predictive value of the underlying pathology in comparison to structural networks where tau pathology predicted connectivity dysfunction in PSP subjects, but no relationship was found in CBD. The authors reasoned that it could be due to the ceiling effect of neurodegeneration in their patient cohort, suggesting that disease stage matters when assessing the association between pathology and functional connectivity. Nevertheless, these results suggest a potential difference in patho-mechanism between the two 4R-tauopathies and that atrophy measure is likely more robust as an *in vivo* biomarker of tau burden and neurodegeneration than functional measures in 4R-tauopathies.

A more recent study (Faye Carlos et al., 2022) further probed the relationship between region-specific volume and tau type-specific burden in each brain region of 4R-tauopathies (PSP and CBD combined). They found that total tau burden was associated with atrophy only in subcortical, but not cortical regions. There was also no relationship between atrophy and NFT burden, but glial tau burden was associated with atrophy in subcortical regions such as the midbrain and subthalamic nucleus. These suggest that the atrophy-tau relationship in 4R tauopathies may be primarily driven by glial pathology due to their abundance across subcortical regions.

The differences in the results between Spina et al (2019) and Carlos et al (2022) in the relationship between neuronal tau and atrophy may stem from the fact that Spina et al included more measures of neuronal tau burden and studied PSP and CBD separately. Furthermore, for white matter, total tau burden in the frontal gyrus and premotor cortex was associated with increased mean diffusivity in the underlying white matter, indicating cellular degeneration. Taken together, these results revealed that the atrophy-tau relationship in 4R-tauopathies may be region and cell type specific. All in all, the investigations thus far have demonstrated that the relationship between neuroimaging measures and tau burden differs between tauopathies. It appears that the imaging-pathology relationship depends on the

severity of the pathology, in which the vulnerability pattern is determined by cell type and anatomical location.

### 1.6.2. Investigating patho-mechanism

Imaging-pathology studies not only help validate *in vivo* biomarkers, but also provide additional insights into the complex interaction between microscopic and macroscopic levels of the brain during disease progression. Despite the rarity of these studies, they are useful in demonstrating the impact of cellular pathology on brain networks and cognition. (Pasquini et al., 2020) investigated the interplay between cellular, brain network and behavioural response in bvFTD. They found that the pathobiology of specific neuron types in layer five of grey matter did not directly cause social-emotional deficits in bvFTD but resulted in the degeneration of the salience network regions which led to behavioural deficits. This finding shows that the involvement of a specific neuronal morphotype embedded in the brain network was critical for cognitive function, in this case empathy.

The relationship between brain networks and tau type-specific pathology can reveal cell type-specific involvement during pathogenesis and suggest new models for understanding the progression of tau neuropathology across the brain. For example, Carlos et al (2022) observed a strong negative association between astroglial tau burden and neurodegeneration in the striatum of 4R-tauopathies. This contrasts with the general trend observed in other subcortical regions which showed positive associations. The author proposed two potential explanations where one concerns the protective role of astrocytes and the other concerns the increase in cellularity due to gliosis. Further investigations are required to untangle this unexpected finding which could reveal novel or supportive insights for the role of striatum and astrocyte during the pathogenesis.

Overall, neuroimaging-pathology studies thus far have demonstrated that they can be used to expand a pool of possible research questions. Direct and specific evidence in assessing tau spread hypothesis need not be restricted to molecular research alone, and that it is now possible to assess the interactions between components from different level of abstractions within the disrupted system which can generate new hypotheses for further investigation. Nevertheless, this is simply the beginning and more significant questions remain in

understanding the impact of neuropathology on specific brain structures and networks across tauopathies.

## 1.7. PSP as a potential disease model

As previously reviewed, tauopathies is a class of neurodegenerative diseases associated with tau neuropathology. To begin investigating abnormal tau using a relatively new approach such as digital pathology, a reasonable first step would be to study a disease with minimal complications. In this thesis, I have chosen to focus on a primary tauopathy where tau can be studied in isolation, unlike Alzheimer's disease where beta-amyloid is also involved, or Parkinson's disease with alpha-synuclein. PSP is the central focus of this thesis as it has a well-defined staging scheme and tau type-specific aggregates are relatively distinct, essential for a proof-of-concept investigation and more.

PSP is an example of a prototypical primary tauopathy (Höglinger et al., 2017, 2018) as it enables the study of tau protein in isolation with clear stages of neuropathology (Kovacs et al., 2020). Classically, PSP is classified as an atypical parkinsonian disorder and is estimated to affect 5-17 in 100,000 people (Elble, 2021). Diagnosis is typically 3-4 years after symptom onset (Höglinger et al., 2017), and life expectancy is approximately 5-7 years (Armstrong et al., 2014). Based on the PSP diagnostic criteria (Höglinger et al., 2017), the core clinical features of PSP include vertical gaze palsy, postural instability, akinesia, and cognitive dysfunction. PSP patients can have varied clinical presentations, with the most common subtype being PSP Richardson syndrome (PSP-RS) which is defined by postural instability and falls with vertical ocular motor dysfunction. Other subtypes include PSP with ocular motor dysfunction (*PSP-OM*), postural instability (*PSP-PI*), parkinsonism (PSP-P), frontal lobe cognitive or behavioural presentations (*PSP-F*), progressive gait freezing (*PSP-PGF*), cortical basal syndrome (*PSP-CBS*), primary lateral sclerosis (PSP-PLS), cerebellar ataxia (PSP-C) and speech disorder (*PSP-SL*).

PSP is characterised pathologically by 4R tau, with pathological hallmarks of neurofibrillary tangles (NFT), neuropil threads, coiled bodies (CB), and tufted astrocytes (TA). Pathological tau is predominantly found in subcortical structures but appears in the neocortex as the

disease progresses. The differences in the distribution of neuropathology explains the multiple PSP subtypes.

Despite the clinical heterogeneity, there are key brain regions in PSP that dictate disease stage (Kovacs et al., 2020). These include highly affected regions, of the subthalamic nucleus, globus pallidus, striatum, and cerebellar dentate nucleus; modestly affected regions of the and frontal cortex; and the occipital cortex affected only in late disease. The ability to examine these regions makes PSP a prime candidate for creating a disease model in improving the understanding of tau pathology and its relationship to neuroimaging biomarkers.

## 1.8. Research objectives

In this thesis, I aimed to develop a digital pathology pipeline to better quantify cellular changes including cell loss and tau protein aggregates and understand their relationship to *in vivo* MRI biomarkers. The main goal of this pipeline is to contribute to research by quantifying detailed measurements at scale with minimal processing time, essential for complex and robust analyses for understanding abnormal tau protein better. I focus on PSP, a primary tauopathy characterised by the accumulation of 4-repeat tau in neuronal and glial cells.

*In chapter 3 (page 58),* I aimed to create a pipeline using a supervised machine learning algorithm to quantify total and tau positive cell type-specific density. Specifically, I aim to:

- Design a pipeline that will perform cell classification and quantify tau positive and negative cells for neurons, astrocytes and oligodendrocytes in control and PSP *post mortem* samples.
- The pipeline should be applicable across key brain regions in PSP pathology.
- The pipeline should have high performance in the held-out test set and yield quantification results that are consistent with the wider literature on slides beyond the held-out test set where there is no ground truth.
- Compare different approaches in the pipeline workflow to ensure the resulting pipeline is optimal in terms of accuracy, interpretability, and adaptability.

Such a task has proven to be extremely challenging, and I demonstrated that using supervised probabilistic machine learning models with cellular features was insufficient to accurately identify cell types in the context of PSP.

*In chapter 4 (page 100),* I adapted the pipeline from project 1, using a probabilistic random forest approach, to identify tau type-specific aggregates in PSP *post mortem* samples. Specifically, a successful pipeline should:

- Be able to quantify density of neurofibrillary tangle, tufted astrocyte, coiled bodies, and tau fragments.
- Be applicable to key brain regions in the current PSP staging system and beyond the staging system.
- Have high classification performance that is comparable to expert neuropathologists in the held-out test set.

*In chapter 5 (page 124),* I set out to apply the tau quantification pipeline on novel slides from PSP and control donors, beyond the held-out test set. Specifically, I aimed to:

- Investigate digitally quantified tau distribution patterns and the relationship between tau aggregate types across brain regions.
- Validate digitally quantified tau burden against the current PSP staging system.
- Investigate the relationship between digitally quantified tau burden, PSP stage and clinical severity in PSP donors.
- Compare digitally quantified tau burden from PSP to control donors to demonstrate that the pipeline can quantify tau burden beyond PSP *post mortem* samples.

*In chapter 6 (page 149),* I focused on validating whether structural and functional MRI measures can be used as *in vivo* biomarkers of tau burden in PSP. To do this, I investigated:

- The relationship between structural atrophy and tau burden (both total and tau type-specific) across brain regions.
- Similarly, the relationship between functional network graph metrics and tau burden (both total and tau type-specific) across brain regions.

## 1.9. Summary and contribution of thesis

Overall, neuroimaging measures are increasingly useful as *in vivo* biomarkers of neurodegenerative diseases. They were previously used to only rule out alternative causes of cognitive impairments but now are included in various diagnostic criteria for tauopathies. However, they still have limited translational value as their specificity and sensitivity as biomarkers require more thorough validation. Tau is a key protein in tauopathies where specific pattern of tau aggregation in an affected cell type and brain region is used to distinguish between different tauopathies and determine disease severity. It is important then to investigate the relationship between *in vivo* imaging measures and tau burden for tracking disease progression, monitoring prognosis and differential diagnosis. PSP is a suitable disease as it is a primary tauopathy which enables the study of tau protein in isolation and has a well-defined pathological staging system.

Neuropathological examination of *post mortem* data is the gold standard approach for definite diagnosis of tauopathies as tau distribution pattern can characterised. This makes it important to derive tau burden from such data for validating *in vivo* neuroimaging markers. However, neuropathological examination is a semi-quantitative approach with limitations. It is a subjective and time-consuming process that requires high level of expertise, making it prone to both intra- and inter-rater variability. With this, limited brain regions can be analysed and a simple ordinal measure of pathological stage limits more complex research questions and analyses to be possible. Automated pipeline using machine learning algorithms can be applied to move towards a more objective, detailed, and scalable solutions for pathological assessment. However, there also remains limitations to be addressed such as validation, interpretability, and standardisation.

In conclusion, this thesis attempts to firstly develop a digital pathology pipeline for quantifying PSP-related pathology across key brain regions in the PSP staging system and more. The pipeline will be validated on a held-out test set and all available novel slides where quantification results would be compared against the PSP staging scheme. With the quantification results, the relationship between tau aggregate types, comparison between PSP and control tau burden, and the relationship between neuropathological and clinical severity

would be investigated. Finally, digitally quantified tau burden will be used to validate structural and functional MRI measures as *in vivo* biomarkers of tau burden and provide additional insights for the role of tau in the pathogenesis of PSP.

# Chapter 2: Core methods

In this thesis, I will be using cutting edge and established methods applied to histopathology and neuroimaging data. These methods form the basis of work in the rest of the chapters. I will firstly outline the digital pathology pipeline which is the core method used to perform cell and tau type-specific classification. Bayesian statistics will also be outlined as it is the main statistical method used throughout the chapters to investigate the relationship between tau burden and other measures such as cognitive score and neuroimaging measures.

This work was performed in collaboration with Sanne Kaalund (post-doctoral researcher with James Rowe and the Cambridge brain bank) who extracted brain slices, arranged slide staining and scanning, and led the manual annotation of cell types; Eric Hidari (post-doctoral researcher with James Rowe) who assisted in developing a pipeline using QuPath software for brain slide segmentation.

## 2.1. Digital pathology pipeline

The core steps in the digital pathology pipeline include data preparation, feature selection, supervised machine learning algorithms and model development.

## 2.2. Data preparation

### 2.2.1. Tissue processing and immunohistochemistry

*Post mortem* tissue with a clinical and pathological diagnosis of Progressive Supranuclear Palsy (PSP) were obtained from the Cambridge Brain Bank under the Neuropathology Research in Dementia (NERD) study with ethical approval from the Wales 6 Research Ethics Committee. For each participant, the brain was removed at autopsy for *post mortem* analysis and the left cerebral hemisphere and left cerebellum were fixed in 10% neutral buffered formalin for two to three weeks. Fixed brain tissues were embedded in paraffin and cut at 10 μm on a rotary microtome and mounted on microscope slides. Sections were deparaffinised in xylene and dehydrated through 2 troughs of 99% Industrial Denatured Alcohol (IDA) and

rinsed in running water. Sections were incubated in formic acid for antigen retrieval and endogenous peroxidase blocked with a 4% peroxide solution. Sections were then incubated in phosphate buffered saline and blocked in 5% normal rabbit serum before incubating with primary antibody. Slides were incubated with AT8 (1:500, MN1020, Thermo Scientific, USA) and washed in phosphate buffered saline and incubated with the secondary antibody, Polyclonal Rabbit Anti-Mouse Immunoglobulins/Biotinylated (E0354, Dako, Denmark). The signal was amplified with the ABC kit (Vectastain), washed in phosphate buffered saline, then the chromogen is 3,3'-diaminobenzidine (DAB) was applied to visualize pathological tau as a brown reaction product. Counter-staining was performed using haematoxylin to visualise cell nuclei as blue reaction products. Slide images were acquired by an Aperio AT2 whole slide scanner (Leica) at 40x magnification with a resolution of 0.2528 μm / pixel.

### 2.2.2. Image pre-processing

All pre-processing steps (Figure 2.1) were carried out in QuPath (version 0.4.3) software (Bankhead et al., 2017). First, color deconvolution was applied to all scanned bright-field (H-DAB) whole slide images to digitally separate stains into three different channels: the DAB channel for hyperphosphorylated tau, the hematoxylin channel for cell nuclei and a residual channel. Slides were then manually inspected to remove obvious artefacts such as DAB artefacts, de-focused regions, folded tissue, air bubbles and other confounding objects. Brain tissue was separated from the background and segmented into respective regions; for cortical regions, a semi-automated grey and white matter segmentation was carried out using the simple tissue detection tool, followed by the wand tool to manually fine-edit the segmentation. For basal ganglia regions, putamen, globus pallidus and subthalamic nucleus were manually segmented by an experienced neuropathologist (SSK). The dentate nucleus was segmented from the cerebellum slide by a trained expert (TP).

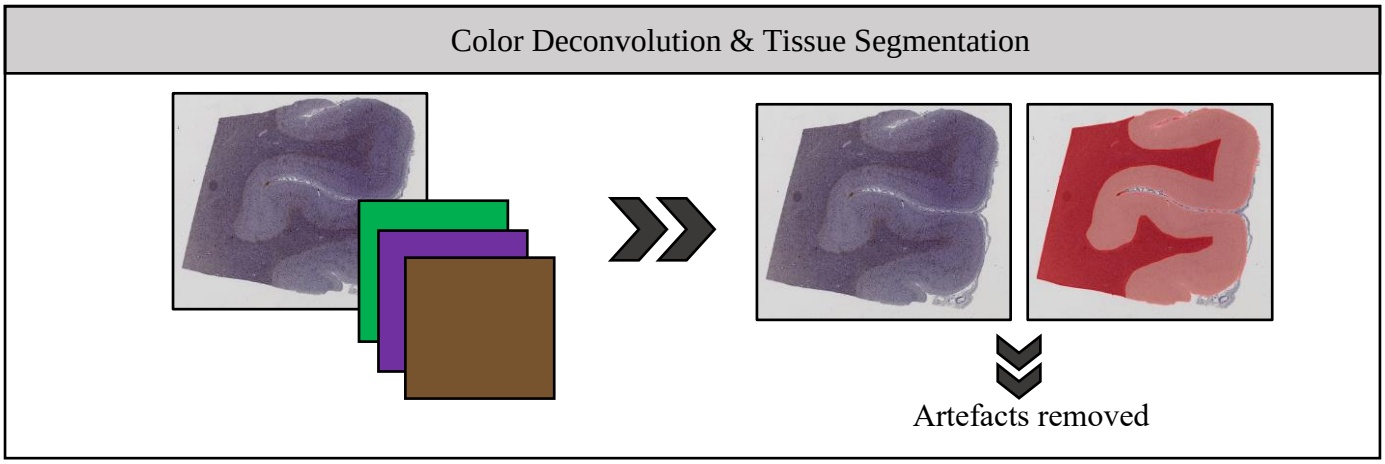



Figure 2.1: A whole slide image is imported into QuPath which undergoes colour deconvolution and segmentation of region of interest. Artefacts are manually removed at this step.

## 2.3. Feature selection

In a dataset, the rows contain data points, and the columns contain features, where they can be of various types such as categorical or numerical. Feature selection is a process of selecting only informative or relevant features in a dataset for a machine learning model to learn and perform a specific task optimally. This step is crucial as it helps reduce uninformative and redundant features to ensure a parsimonious model with optimal performance and minimal computational complexity.

### 2.3.1. Recursive feature elimination

Recursive feature elimination (RFE) is a feature selection algorithm (or set of rules) that fits a model to the data with all available features and removes the least important feature, until a specified number of feature is reached (Guyon et al., 2002). Starting with fitting a chosen machine learning model such as a support vector machine with a full feature set, 1) feature importance is calculated for all features, 2) the lowest ranked feature is removed, 3) the remaining features are refitted to the model. Step 1 to step 3 are repeated until a chosen number of features remain. RFE is a flexible approach that can be wrapped around any supervised learning model, works well with complex datasets in selecting important features and can tackle interactions between features. However, it is computationally expensive as it is a recursive approach and may have issues with highly correlated and noisy datasets.

### 2.3.2. Principal component analysis

Principal component analysis (PCA) is a dimensionality reduction approach where an original feature set is decomposed to a lower dimensional space, resulting in a smaller number of components that still capture most of the variance in the dataset (Pearson, 1901). This means PCA is suitable when the feature set is large and when features are highly correlated as PCA can lessen the impact of multicollinearity by identifying the most important features or components. This can improve visualisation and speed up the processing time for machine learning algorithms. However, it can be difficult to interpret the components as they are transformed from the original feature space and may not be suitable when there is a non-linear relationship between features.

## 2.4. Supervised machine learning algorithms

In supervised learning, machines learn appropriate actions from examples given to them (Cunningham et al., 2008). They are provided with inputs and correct outputs, where they need to identify patterns in the data to make correct predictions. Supervised algorithms can be used for classification task where they learn to predict the category of a new instance based on previous observations (Guyon et al., 2008). There is no universal best classifier for every problem as every dataset is different (Wolpert & Macready, 1997). In this thesis, two machine learning algorithms are used: support vector machines and random forests. Further details of how these are applied are included in each chapter, however there are specific considerations in the choice of machine learning model that I will discuss here.

### 2.4.1. Class imbalance issue

Class imbalance is an inherent attribute of most real-world datasets which poses issues to most machine learning algorithms (Luo et al., 2019; More & Rana, 2017). For a classification task, algorithms usually assume that the number of instances in each class or category are equal. Therefore, the predictive performance of the algorithm trained on an imbalanced dataset where the number of instances in each class is not equal will be poor. The algorithm becomes biased towards the majority class with higher proportion of instances in the dataset than other classes, and against the minority class which has smaller proportion of instances. This results in the over-classification of the majority class and under-classification of the

minority class. This issue becomes particularly important in tasks where the minority class is more important or of more interest than the majority class. In this thesis, 3 methods to tackle class imbalance are explored: cost-sensitive learning, a re-sampling strategy, and the threshold-moving approach. *Cost sensitive learning* involves changing the weight or penalty parameters of the algorithm (C. X. Ling & Sheng, 2010), by for example, introducing the *class weight* parameter to penalise classification mistake more in the smaller than the larger class. A *re-sampling strategy* directly changes the class distribution by re-sampling the dataset (Chen, 2011). In general, machine learning algorithms can predict probability or scores of class membership. The scores would then be 'thresholded' to map class scores into class labels. For a binary classification task, the threshold could be 0.5 for both classes (totalling 1). An object would be mapped to a specific class label when the class-specific score is equal to or greater than the class-specific threshold. With severe class imbalance, the default class-specific thresholds could lead to suboptimal results. The *threshold-moving method* can be used to tune the class-specific thresholds. I have chosen to tune for optimal thresholds by using the precision-recall curves as it is suitable for when there is moderate to large class imbalance (Davis & Goadrich, 2006).

### 2.4.2. Support vector machine

A support vector machine (SVM) is an algorithm that constructs a hyperplane to separate two classes of data (Cortes & Vapnik, 1995). The hyperplane is constructed to ensure the distance (or margin) between itself, and data points is maximised for minimal generalisation error and maximal confidence. SVM can be applied to both linearly and non-linearly separable data, depending on the choice of kernel functions for data transformation. A common first choice of non-linear kernel is the radial basis function (RBF) which is the least complex kernel and is relatively similar to a linear kernel with an additional hyper-parameter (Hsu et al., 2003). A linear SVM (L-SVM) has 1 hyperparameter; the penalty regularisation parameter C, while an SVM with RBF kernel (RBF-SVM) has an additional regularisation parameter $\gamma$. Both parameters regulate the complexity of the hyperplane, where C controls the error and $\gamma$ controls the complexity and flexibility of the hyperplane.

The standard SVM can be adapted to tackle with class imbalance using the cost-sensitive learning approach by introducing the *class weight* parameter and the threshold-moving method to tune class probability estimates.

### 2.4.3. Random forests

A random forest (RF) is a tree-based ensemble algorithm, meaning that the final prediction is a result of a number of separately run predictions (Breiman, 2001). The data is re-sampled to create many bootstrap (smaller) datasets. A decision tree is created for each random subset of each bootstrap dataset. This is important as it de-correlates the decision trees, improving the ensemble decision. The RF model considers class prediction voting from all trees in the forest and finally outputs a class prediction with the majority vote. A random forest algorithm has many hyperparameters but in general, they relate to determining the number of trees and leaves (nodes) in the forest, number of features to consider at each node, and re-sampling strategy.

In this thesis, I have chosen to apply cost-sensitive learning and the re-sampling technique to the standard random forest. Therefore, a balanced random forest which randomly under-samples the majority class in each bootstrap, making the data balanced (Chen, 2011) is used with a class weight parameter.

As a random forest classifier makes a final class prediction based on majority voting, it operates under the assumption that each class has an equal likelihood, or threshold, of occurring. The threshold can be adjusted to further tackle severe class imbalance issue using a threshold-moving technique (Lipton et al., 2014; X. Zhang et al., 2020; Zou et al., 2016). This is especially relevant for tau burden classification as their relative proportions are different in cortical and subcortical structures (Kovacs et al., 2020).

## 2.5. Model development

### 2.5.1. Area under precision-recall curve as a performance metric

To evaluate classification performance of a probabilistic classifier, area under the precision-recall curve (PR-AUC) can be used. A precision-recall curve is a plot with *precision* on the y-axis and *recall* on the x-axis. Precision (or positive predictive value) is the proportion of relevant and retrieved items amongst all the retrieved items. Recall (or sensitivity) is the proportion of relevant and retrieved items amongst all the relevant items. To create a precision-recall curve using The Sci-kit learn library (version 0.24.1) in Python (Pedregosa et al., 2011), predicted class scores and true class labels for objects are fed to *precision_recall_curve()* function to compute precision-recall pairs for different (probability or score) thresholds (from 0.0 to 1.0). This function is designed for a binary classification task, but cell or tau classification is a multi-class classification task. One-vs-rest approach is used to split up the multi-class classification task into multiple binary classification problems where a class is compared against all other classes. With this, each class has its own precision-recall curve and *auc()* function was applied to calculate area under the curves. Mean value of all areas across all classes was computed and used as a performance metric in the model development stage such as tuning hyper-parameters of the machine learning algorithm.

### 2.5.2. Sample size check

For my specific cell or tau classification task, no ground truth is available to indicate the training sample size required for an optimal classification performance. This is crucial as too little data results in suboptimal performance and low generalisability, while too much data means unnecessary extra time on data preparation and computational complexity. Therefore, to ensure sufficient objects were annotated for training the machine learning algorithm, I compared classification performance obtained from using smaller portions of the dataset to the full dataset. With this, 10% of the dataset was first randomly sampled while preserving the class ratio to train the classifier using 10-fold cross validation to obtain PR-AUC. This was repeated with an additional 10% of the dataset each time, until 100% of the dataset was used. PR-AUC from the ten sampling points were plotted for visual inspection where PR-

AUC should initially increase with training size to indicate that the algorithm learns better as more data is provided. When the plot plateaus, this indicates that extra training data does not further improve classification performance, suggesting that sufficient data has been used to train the algorithm for optimal results.

### 2.5.3. Hyper-parameter tuning

The Sci-kit learn (version 0.24.1) (Pedregosa et al., 2011) and Imbalanced-learn (Lemaître et al., 2017) libraries in Python (version 3.8.1) were used to implement all digital pathology pipelines. Annotated objects from all slides were pooled together, yielding a dataset. The dataset was then standardised (*mean = 0, SD =1*) and 10-fold stratified cross validation was used to train the classifiers.

Feature selection approaches have the following parameters, *n_features_to_select=[28, 30, 34, 36, 38, 40, 42, 44]* (number of features to select for RFE) *and n_components_to_select = [0.95, 0.96, 0.97, 0.98, 0.99] (select the number of components such that the amount of variance that needs to be explained is greater than n_components_to_select).*

Hyperparameters define the architecture of a machine learning model and are chosen before training. During training, hyper-parameter tuning is carried out to determine the optimal set of model parameters. Hyperparameters of SVM classifiers were tuned using grid-search with $C = [10^{-1}, 10^{0}, 10^{1}, 10^{2}]$ and $\gamma = [10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}]$ when RBF kernel was used. Hyper-parameters of the balanced random forest were tuned using a random-search with the following parameter space: *n_estimators = [100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]* (number of trees in the forest), *max_features =[0.2, 0.4, 0.6, 0.8, 1]* (number of features to consider for best split), *max_depth= [5, 10, 15, 20, None]* (maximum depth of the tree), *min_samples_split = [2, 5, 10]* (minimum samples required to split further), *min_samples_leaf = [1, 2, 4]* (minimum samples required to be a leaf node), *sampling_strategy = ['auto', 'all', 'not majority', 'majority']* (sampling strategy to sample the dataset), *max_samples = [0.25, 0.5, 0.75, None]* (number of bootstrap samples to draw to train each base estimator), *class_weight = ['balanced']* (weight or importance associated with the classes). Both classifiers were optimised based on the mean area under the 4 precision-

recall curves (PR-AUC) using a one-vs-rest approach (class 1 vs rest, class 2 vs rest, class 3 vs rest, class 4 vs rest).

### 2.5.4. Class-specific threshold tuning

Using the hyper-parameters found, optimal class-specific thresholds were tuned to tackle class imbalance. Predicted class scores and true class labels were used to re-compute precision-recall curve for each class using a one-vs-rest approach. Class-specific threshold was optimised using F1-score calculated from the precision-recall pairs on the class-specific curve. F1-score was computed as follows.

$$F1 = 2x\frac{Precision\ x\ Recall}{Precision + Recall}$$

After obtaining class-specific thresholds, class probabilities for each object were thresholded to obtain the predicted class label. Brain regions with similar tau or cell morphology and distribution were grouped together where class-specific thresholds were tuned separately for each brain grouping.

If an object's class probability passed the class-specific threshold, an object would be labelled as the corresponding class. To mirror human classification of tau or cell objects, I assessed the ambiguity of tau or cell object classification. If more than one class or no class passed the class-specific threshold, the object was labelled as 'Ambiguous' and discarded from further analyses.

### 2.5.5. Classification performance

After classification, the precision, recall, macro F1-score and confusion matrix of the model were collected. Macro F1-score was calculated using the following formulae:

$$MacroF1 = \frac{\sum(F1 - scores)\ all\ classes}{number\ of\ classes}$$

The model was then applied to the held-out test set to evaluate its performance generalisability. Finally, the optimised model was applied to the remaining novel slides for classification and quantification and further analyses.

## 2.6. Bayesian statistics

Bayesian analysis was chosen due to several advantages (Kruschke et al., 2012), in particular, it enables the calculation of posterior probability distributions showing the uncertainty of the regression coefficient estimates and that sample size does not affect the inference method. In addition, null hypothesis could also be rejected or accepted which is of interest in this study. However, it is more computationally intensive than frequentist statistics and the results may be sensitive to the priors used. Here, I will describe the core of Bayesian statistics and regression, configurations of Bayesian regression models used in this thesis and steps taken to address potential issues.

Bayesian statistics can be described as the reallocation of credibility (or probability) across possibilities (Kruschke & Liddell, 2018b). With some data to be explained, we have an existing set of candidate explanations. Before observing any new data, these explanations have 'prior credibility' of them being the best explanations. After observing the new data, we re-allocate the credibility of the candidate explanations towards the ones that better explain the data, and away from those that do not. This idea can be formalised as Bayes' rule, where $C_k$ is class 'k' and x is data:

$$P(C_k|x) = \frac{P(C_k)P(x|C_k)}{P(x)}$$

$$\propto P(C_k)P(x|C_k)$$

$P(C_k|x)$ is the posterior probability; the probability of class 'k' given the data and concerns the re-allocation of distribution of probability after observing the new observation. $P(C_k)$ is the prior probability; the probability of each possibility before seeing the data. $P(x|C_k)$ is the likelihood; the probability of observed data given any candidate values of the parameters. P(x) is the probability of the data or evidence.

### 2.6.1. Bayesian regression

To investigate the relationship between independent and dependent variables, the *brms* package in R (Bürkner, 2017, 2018, 2021) was used to construct Bayesian linear regression models. Multiple models were constructed to test various hypotheses across the projects where Bayes' factor (BF) was used for model comparison. To have sufficient evidence in favor of an alternative model against a baseline model, BF has to be greater than 3 (Kruschke, 2011). In the final model, the strength of the regression coefficient was assessed using the region of practical equivalence (ROPE) which is a range of values that are equivalent to the null value for practical purposes (Kruschke & Liddell, 2018b). If no prior information is available to inform the ROPE, by convention it is constructed as a range of values +/- 0.1 of the standard deviation of a standardized parameter (dependent variable) (Kruschke, 2018). If 95% of the credible interval (Crl) of the regression coefficient falls completely within ROPE, then the effect of the parameter is equivalent to the null value of practice purpose (Kruschke, 2011; Kruschke & Liddell, 2018b). If the 95% Crl falls completely outside ROPE, the ROPE value (but not the entire interval) is rejected. If 95% Crl overlaps with ROPE, the decision remains undecided.

Model configuration was the same for all models (warmup = 10000, iteration = 20000). All models went through prior and posterior predictive checks to ensure that the configurations were valid. All models converged with no divergences or diagnostic warnings where $\hat{R}$ convergence values were all ~ 1.00 (by convention these should be < 1.05). Sensitivity analysis of prior was supplemented for all final models to illustrate the choice of prior on posterior estimates (See Appendix)

# Chapter 3: Total and tau positive cell type-specific quantification

## 3.1. Introduction

Many neurodegenerative diseases are characterised by abnormal protein accumulation within neurons and glia, accompanied by neuronal loss and gliosis (Irwin, 2016; Kovacs et al., 2020). Understanding the severity and distribution of this protein pathology is key to investigate the aetiology, understand disease heterogeneity, model disease progression, and to design molecular-targeted disease-modifying therapies. Hyperphosphorylated and misfolded aggregates of tau accumulate in common and rare neurodegenerative diseases, including AD, FTD, and PSP. Such tau pathology is related to neuronal loss (Giannakopoulos et al., 2003), grey matter atrophy (Whitwell et al., 2008) and clinical severity (Bejanin et al., 2017; Cho et al., 2016). It is therefore interesting to quantify and investigate the relationship between total and tau positive cell type-specific density in the same sample to investigate tau distribution pattern and better understand the impact of tau on neurodegeneration.

To quantify both total and tau positive cell type-specific density, the first step is to perform cell classification, then select a tau threshold to detect tau positive cells. Cell classification using automated pipelines have many applications in biological science and medicine that relate to deciphering various intricate biological processes (Shifat-E-Rabbi et al., 2020). Applications include, for example, understanding the effect of genes (Conrad & Gerlich, 2010) or drugs on cell cultures in screening experiments (Murali et al., 2019; Xu et al., 2008) and detecting cancerous cells for diagnosis and prognosis (Kantara et al., 2015; Oei et al., 2019). Various algorithms have been utilised, notably numerical feature engineering and neural network algorithms (Shifat-E-Rabbi et al., 2020). Numerical feature engineering involves using expert's knowledge to define important features for cell classification such as staining agents (Solorzano et al., 2021) or cytoskeleton features (Oei et al., 2019). Neural network algorithms, which can work directly with images without feature extraction, have been shown to achieve high classification performance with a large amount of training data (Gao et al., 2017; Phan et al., 2016) . Despite a large number of cell classification studies, to my knowledge, there has been only one study that tried to classify neuronal and glial cells using an automated pipeline

(Solorzano et al., 2021). The classification performance was relatively high at 83% and could be pushed to 95% when discarding low-confidence classifications. Still, this study used novel staining techniques with many cell-specific molecular markers.

For automated cell classification pipeline to be widely adopted, data pre-processing should be as simple as possible. No study has yet attempted to perform neuronal and glial cell classification using standard staining agents from routine autopsy assessment, especially in a PSP cohort. Studies using automated pipelines to investigate tau pathological hallmarks in tauopathies have not directly quantified cell density (Koga et al., 2021, 2022). This motivates the development of a cell quantification pipeline that can also detect tau positive cells to aid further investigation of the tau distribution pattern and the relationship between tau aggregates and neuronal loss in human *post mortem* data.

## 3.2. The present study

I attempted to create a cell quantification pipeline using a supervised machine learning algorithm for the cortex and basal ganglia of PSP participants. My objective was to quantify cell density for neurons, astrocytes, and oligodendrocytes. With an additional step of tau thresholding to detect tau positive cells, I also aimed to investigate the relationship between neuronal loss and tau type-specific positive cell density.

### 3.2.1. Pipeline framework

First, I used a supervised machine learning algorithm to perform neuronal and glial classification on *post mortem* brain slides from PSP and control participants. Both PSP and control slides were included to ensure the pipeline can generalise to both groups. There were 3 main methodological challenges in cell classification across multiple brain regions to address. First, there are not equal numbers of neuronal and glial cells in the brain (von Bartheld et al., 2016; WallÃ¸e et al., 2014), leading to a class imbalance for the machine learning model. Second, the ratio of class imbalance and cell morphology differed between brain regions (Ribeiro et al., 2013; Salvesen et al., 2015; von Bartheld et al., 2016). Third, there is inherent ambiguity in classifying some cells, even for expert neuropathologists (Oei et al., 2019; Scheltens & Rockwood, 2011). I therefore designed our pipeline architecture with these challenges in mind (Figure 3.1, Figure 3.2). Supervised machine learning

algorithm that can tackle class imbalance was chosen. By developing three separate classifiers specific for three different groups of brain regions (non-occipital cortical regions, occipital region, basal ganglia), we were able to optimise the classifier within each region. Finally, we explicitly addressed the challenge of ambiguous classification by optimising thresholds for each cell class and excluding individual objects that met either no class threshold or multiple class thresholds.

### 3.2.2. Comparison between technical approaches

In general, a classification pipeline consists of 4 main steps: 1) data preparation, 2) feature selection, 3) model development and 4) testing on novel datasets. Multiple techniques are available for each step and there is no best configuration for every dataset. In this study, I aimed to use the pipeline framework previously described to experiment with different choices of *feature selection* methods and *machine learning algorithms***.** First, features extracted for cell classification in the current study are closely related, where some may be redundant or introducing noise and impair classification performance. Therefore, I aimed to compare recursive feature elimination (RFE) which uses key features for cell classification to principal component analysis (PCA) which reduces original features into components, reducing multicollinearity. Second, as there is no universal best classifier for every dataset, I aimed to compare 3 widely used machine learning algorithms together: a support vector machine with 1) linear kernel (L-SVM), 2) with radial basis function (RBF-SVM), and 3) balanced random forest (RF).

### 3.2.3. Research objectives

I set out to:

1. Compare classification performances of L-SVM when full feature set was used to when RFE or PCA was used to determine the most suitable feature selection approach given the data.
2. With the chosen feature selection approach, 3 machine learning algorithms are compared to determine the most optimal algorithm for cell classification.

3. The final model will be validated on the held-out test set and is expected to have high classification performance, comparable to training performance and expert neuropathologist.
4. The final model will be applied to all available novel slides from PSP and control donors. With no ground truth available, the quantification results, including glia:neuron ratio (GNR) and cell type-specific density in controls should be consistent with the wider literature.
    - For controls, GNR is hypothesised to be higher in subcortical than cortical regions. GNR in cortical regions may range from 1-3.6, with occipital showing the highest GNR as compared to other cortical regions (Ribeiro et al., 2013; von Bartheld et al., 2016).
    - PSP participants are expected to show lower neuronal density than controls due to neuronal loss. This means PSP participants should also show higher GNR than controls.
    - For PSP participants, neuronal density is hypothesised to be lower in higher than lower PSP pathological stages due to neuronal loss.
5. The pipeline will be used to quantify tau positive cell type-specific density to assess whether the pipeline is robust against tau aggregation. In PSP, tau positive cell density should be higher in basal ganglia than cortical regions.

This work was performed in collaboration with Sanne Kaalund (post-doctoral researcher with James Rowe and the Cambridge brain bank) who extracted brain slices, arranged slide staining and scanning, and led the manual annotation of cell types; Mayen Briggs (a neuropathologist at Cambridge Brain Bank) who annotated cell types for inter-rater agreement; and Eric Hidari (post-doctoral researcher with James Rowe) who assisted in developing a pipeline using QuPath software for nuclear detection and brain slide segmentation.

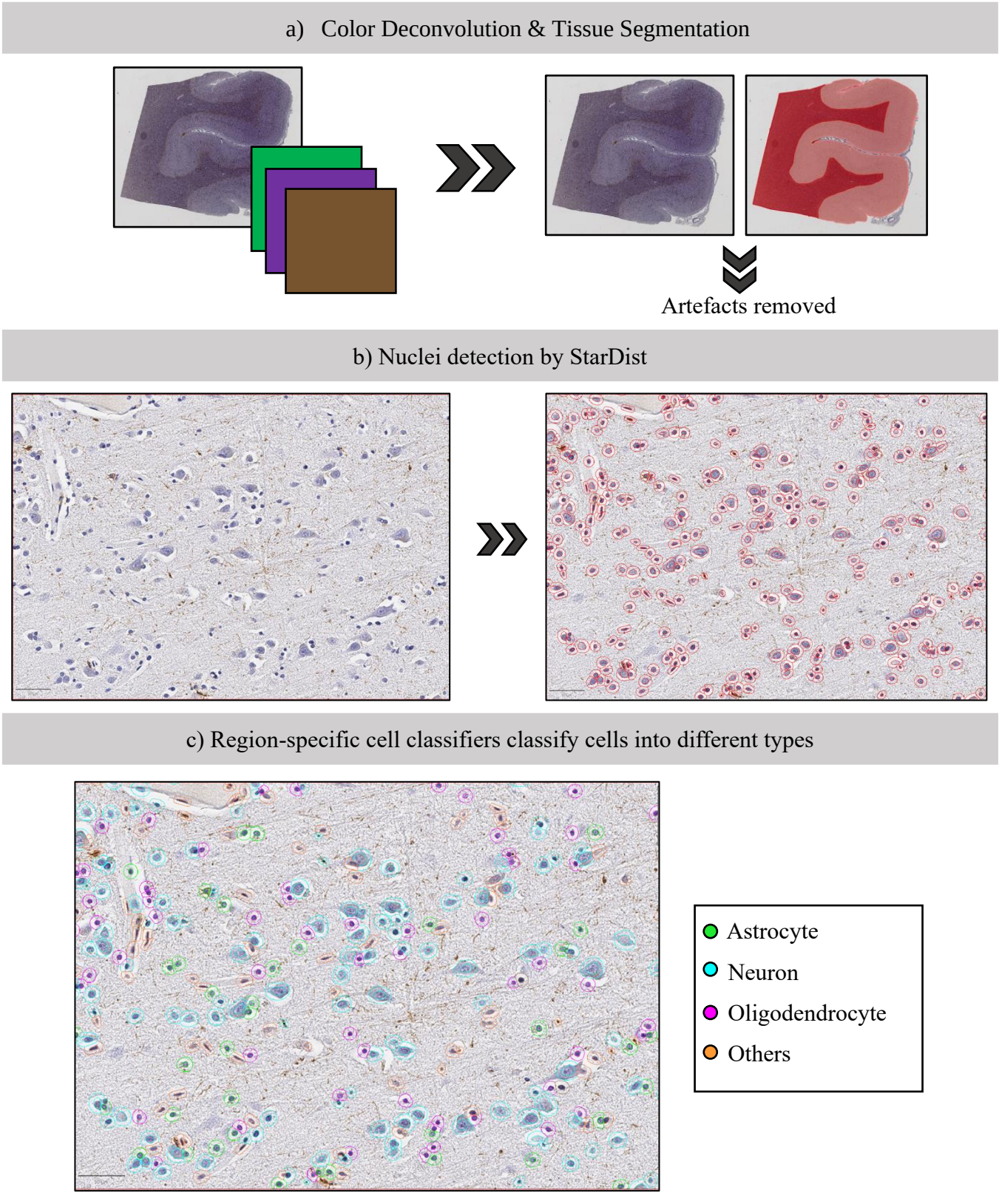



Figure 3.1: Cell classification pipeline overview. a) A whole slide image is imported into QuPath which undergoes colour deconvolution and segmentation of region of interest. Artefacts are manually removed at this step. b) Nuclei detection and cell expansion using StarDist plug-in was applied. Detections with haematoxylin staining in the top 1% were discarded as artefacts. c) Detected objects are fed into the region-specific cell classifiers to separate them into different cell types (astrocyte, neuron, oligodendrocyte, others). Final slide checking should be done to ensure accurate results before subsequent analysis such as cell quantification.

## 3.3. Materials and methods

### 3.3.1. Subject and brain regions

A total of 280 slides were obtained from systematically sampled cortical and subcortical regions from 36 brains donated by patients with a clinical and pathological diagnosis of Progressive Supranuclear Palsy (PSP) and 9 control *post mortem* brains with no known history of neurodegenerative diseases (Table 3.1,Table 3.2). Of these, 28 slides were used for model development (See Table 3.3). From the remaining 252 slides, 10 slides were used as held-out test set (2 slides from each of the following regions: frontal, parietal, temporal, occipital and basal ganglia regions). Of these, 6 slides (frontal, parietal and temporal slides) were independently annotated by 2 pathologists (S.S.K and M.B) to calculate inter-rater reliability. All slides that were not used during model development were used for further statistical analyses (See Table 3.4).

Table 3.1: Demographic, clinical and region-specific severity rating, overall PSP stage and clinical diagnosis of donor participants in the study. Not applicable (N/A) where data is not available in the database.

| Subject | GP | STN | STR | PF | DN | OC | Stage | Clinical diagnosis | Gender | Age at death (years) | Disease duration (years) | Last PSPRS total | PSPRS to death (years) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 2 | 0 | 1 | 0 | 2 | prob. PSP-RS | Female | 76.4 | 8.75 | 63 | 0.32 |
| 2 | 3 | 3 | 2 | 0 | 1 | 0 | 2 | poss. PSP-PGF | Male | 75.2 | 4.62 | 26 | 0.52 |
| 3 | 2 | 2 | 2 | 1 | 1 | 0 | 3 | prob. PSP-RS | Female | 55 | 5.5 | 53 | 0.41 |
| 4 | 2 | 2 | 2 | 1 | 1 | 0 | 3 | prob. PSP-RS | Male | 74.5 | 6 | 45 | 0.67 |
| 5 | 2 | 2 | 2 | 1 | 1 | 0 | 3 | prob. PSP-RS | Female | 84.7 | 5.58 | 59 | 1.25 |
| 6 | 2 | 2 | 2 | 0 | 2 | 0 | N/A | prob. PSP-RS | Female | 73.7 | 16.33 | 51 | 0.56 |
| 7 | 2 | N/A | 2 | 0 | 2 | 0 | N/A | s.o. PSP-P | Male | 74.2 | 8.07 | N/A | N/A |
| 8 | 2 | 2 | 2 | 1 | 2 | 0 | 4 | poss. PSP-CBS | Female | 79.8 | 3.42 | 49 | 0.74 |
| 9 | 2 | 2 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 78.7 | 5.33 | 54 | 0.56 |
| 10 | 2 | 3 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 80.8 | 11.92 | 76 | 1.3 |
| 11 | 2 | 3 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 80.5 | 6.5 | 62 | 0.11 |
| 12 | 2 | 3 | 3 | 1 | 2 | 0 | 4 | prob. PSP-RS | Female | 71.6 | 4.58 | 45 | 0.13 |
| 13 | 3 | N/A | 2 | 1 | 2 | 0 | 4 | s.o. PSP-CBS | Female | 75.2 | 2.83 | N/A | N/A |
| 14 | 3 | 3 | 3 | 3 | 2 | 0 | 4 | poss. PSP-CBS | Male | 78.8 | 5.75 | 43 | 0.69 |
| 15 | 3 | 3 | 2 | 2 | 3 | 0 | 4 | prob. PSP-RS | Male | 77.4 | 6.33 | 55 | 0.74 |
| 16 | 2 | N/A | 2 | 2 | 3 | 0 | 4 | prob. PSP-RS | Male | 64.7 | 5.08 | 38 | 1.09 |
| 17 | 2 | 3 | 2 | 1 |  | 0 | 4 | prob. PSP-RS | Female | 65.3 | 13.92 | 54 | 0.75 |
| 18 | 2 | 2 | 2 | 2 | 2 | 1 | 5 | prob. PSP-RS | Female | 74.7 | 6.58 | 58 | 2.66 |
| 19 | 2 | N/A | 2 | 2 | 2 | 1 | 5 | prob. PSP-RS | Male | 88.5 | 5.08 | 53 | 0.3 |
| 20 | 2 | 3 | 2 | 3 | 2 | 1 | 5 | prob. PSP-RS | Male | 71.5 | 5.17 | 38 | 2.21 |
| 21 | 3 | 3 | 3 | 3 | 2 | 1 | 5 | prob. PSP-RS | Female | 78.1 | 16.75 | 52 | 0.99 |
| 22 | 3 | 3 | 3 | 1 | 3 | 1 | 5 | poss. PSP-CBS | Female | 78.9 | 5 | 48 | 0.18 |
| 23 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Female | 69.9 | 5.33 | 51 | 0.36 |
| 24 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Male | 63 | 8.83 | 62 | 2.44 |
| 25 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Male | 84.6 | 8.75 | 59 | 2.42 |
| 26 | 2 | N/A | 2 | 2 | 3 | 1 | 5 | prob. PSP-RS | Female | 71.9 | 6.17 | 60 | 0.05 |
| 27 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Male | 71 | 5.42 | 43 | 1.89 |
| 28 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | poss. PSP-SL | Female | 78.4 | 8.83 | 72 | 1.94 |
| 29 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Male | 76.2 | 3.87 | 51 | 0.42 |
| 30 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | s.o. PSP-CBS | Male | 73.9 | 4 | N/A | N/A |
| 31 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Female | 84.1 | 4.25 | 67 | 0.78 |
| 32 | 1 | 3 | 1 | 2 | 2 | 2 | N/A | prob. PSP-RS | Male | 72.3 | 4.33 | 38 | 1.25 |
| 33 | 3 | 3 | 2 | 2 | 3 | 2 | 6 | s.o. PSP-CBS | Female | 80.5 | 8.42 | 73 | 0.23 |
| 34 | 3 | 3 | 3 | 3 | 3 | 2 | 6 | prob. PSP-F | Male | 75.4 | 8.42 | 81 | 0.76 |
| 35 | N/A | N/A | N/A | N/A | N/A | N/A | 5 | poss. PSP-CBS | Male | 78 | N/A | N/A | N/A |
| 36 | N/A | N/A | N/A | N/A | N/A | N/A | 3 | prob. PSP-RS | Male | 72 | N/A | N/A | N/A |

Severity rating for each brain region includes 0 = absence, 1 = mild, 2 = moderate, 3 = severe. For clinical diagnosis of PSP participants, *prob.* probable, *poss.* possible, s.o.

suggestive of, *RS* Richardson syndrome, *CBS* predominant corticobasal syndrome, *SL* predominant speech and language disorder. *PF* pre-frontal, *1°M* Primary motor, *PM* Pre-motor, *OC* Occipital, *BG* Basal ganglia, *DN* Dentate nucleus.

Table 3.2: Demographic and Braak stage of control donor participants in the study.

| Subject | Diagnosis | Gender | Age at Death (years) | Braak |
|---|---|---|---|---|
| 1 | Control | Female | 70 | I |
| 2 | Control | Male | 68 | II |
| 3 | Control | Male | 87 | I |
| 4 | Control | Male | 70 | N/A |
| 5 | Control | Female | 72 | II |
| 6 | Control | Female | 84 | III |
| 7 | Control | Male | 82 | III |
| 8 | Control | Male | 74 | II |
| 9 | Control | Female | 70 | II |

Table 3.3: Number of slides from PSP and control, used for model development.

| Region | PSP | Control |
|---|---|---|
| Frontal | 2 | 2 |
| Temporal | 2 | 2 |
| Parietal | 2 | 2 |
| Occipital | 6 | 2 |
| Basal ganglia | 7 | 1 |

Table 3.4: Number of novel slides from PSP and control, not used during model development.

| Region | PSP | Control |
|---|---|---|
| Frontal | 35 | 5 |
| Temporal | 23 | 6 |
| Parietal | 23 | 3 |
| Cingulate | 25 | 1 |
| Primary somatosensory | 25 | 0 |
| Primary motor | 23 | 0 |
| Pre-motor | 24 | 0 |
| Occipital | 29 | 2 |
| Basal ganglia | 27 | 1 |

### 3.3.2. Nuclei detection and feature extraction

Table 3.5: Features from each category were extracted from detected cells, including cellular detection probability, cell morphology, haematoxylin staining intensity and number of nearest neighbours by distance.

| Number of features | Category | Feature |
|---|---|---|
| 1 | Cellular detection probability | Probability |
| 12 | Nucleus, cell morphology | Area (μm$^2$)<br>Length (μm)<br>Circularity<br>Solidity<br>Max diameter (μm)<br>Min diameter (μm) |
| 1 | Nucleus/cell area ratio | Nucleus/Cell area ratio |
| 20 | Haematoxylin staining intensity in nucleus, cytoplasm, membrane, cell | Mean<br>Median<br>Min<br>Max<br>Standard deviation |
| 10 | Number of nearest neighbours at each distance location | 10 – 100 μm |

StarDist, a plug-in software for a deep-learning based method was used to perform nuclei detection on all slides (Schmidt et al., 2018). StarDist first detected cell nuclei and expanded around the nuclear boundary to estimate the extent of the cytoplasm, creating an entire cell object. The dilation radius was set to 3 times the radius of the nucleus, or 5 μm, whichever was smaller. The following parameters were chosen: *threshold = 0.5* (prediction threshold), *normalizePercentiles (1, 99)* (percentile normalization), *pixelSize = 0.25* (resolution for detection), *cellExpansion=5.0* (approximate cell size based upon nucleus expansion), *cellConstrainScale=3* (constrain cell expansion using nucleus size), *ignoreCellOverlaps=false* (prevent cells from expanding into one another), *measureShape=true* (add shape measurements), *measureIntensity*=true (Add cell measurements in all compartments), *includeProbability=true* (add detection probability as measurement), *simplify= 1* (Control how polygons are simplified to remove unnecessary vertices). After cell detection, 44 features were extracted from QuPath which were 34 features relating to morphology and staining intensity, and 10 features relating to the number of nearest neighbours within 10-100μm for contextual information of neighbourhood cell density (See Table 3.5).

### 3.3.3. Training set

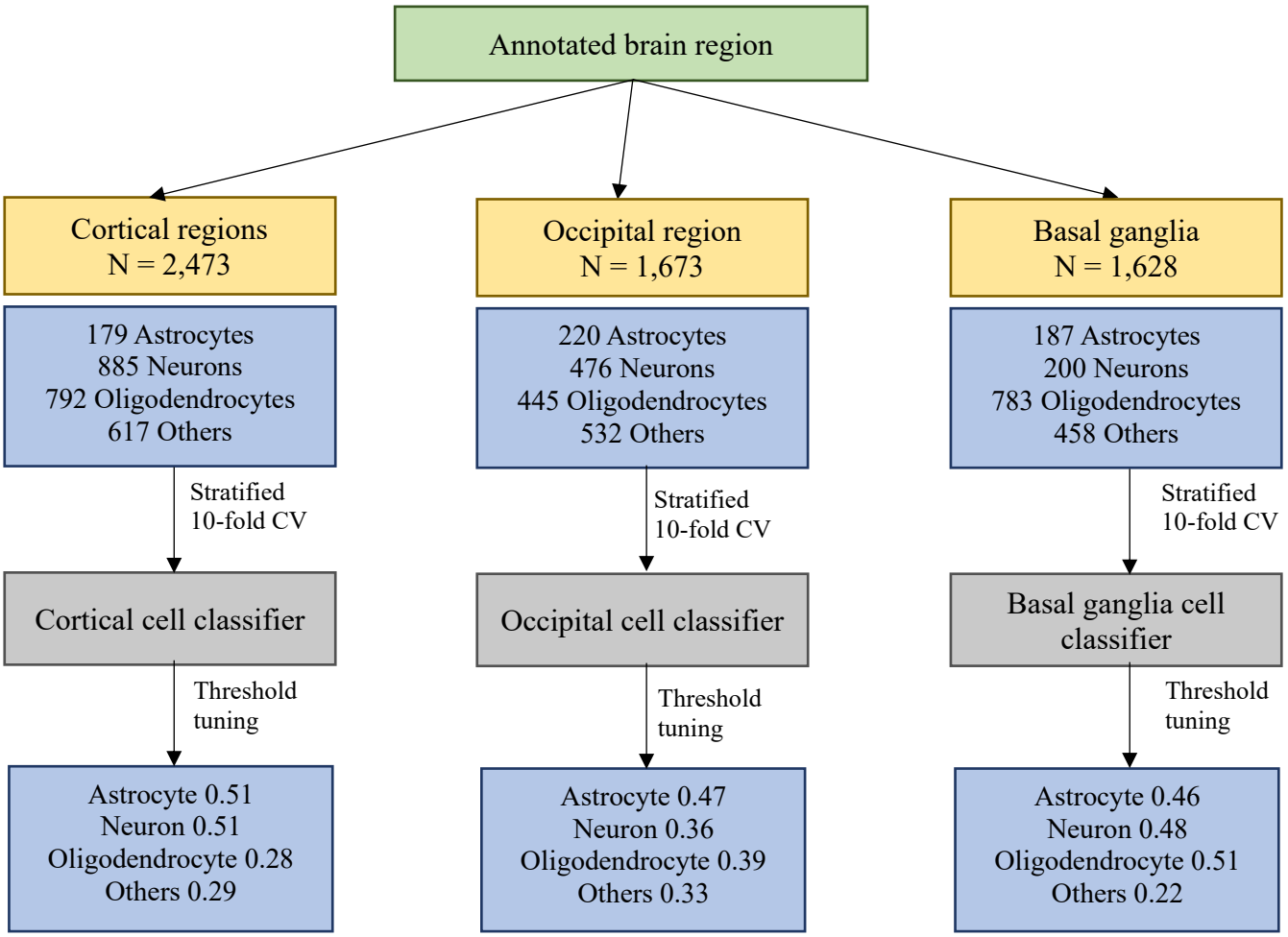
To create the sampling area of cortical training slides, portions of grey matter with no artefact and consistent depth from visual inspection were labelled as ‘cortical ribbon’. A rectangle of equal width was drawn across all layers of grey matter within the cortical ribbon. For basal ganglia, sampling squares were created with equal spacing. All cells were annotated by a single pathology (SSK) as belonging to one of the four classes (‘astrocyte’ (Astro), ‘neuron’ (Neuron), ‘oligodendrocyte’ (Oligo), or ‘others’ (Others). Others class consists of endothelial cells and artefacts on the slide which include ‘fragmented’ cells where cells were partially detected by StarDist and ‘ignored’ cells where StarDist detected out-of-focus cells or detected background as cells. All cells used for training are healthy cells with minimal or no tau aggregation.

For cortical regions excluding the occipital regions, 2473 cells were annotated (179 astrocytes, 885 neurons, 792 oligodendrocytes, 617 others). For occipital cortex, 1673 cells were annotated (220 astrocytes, 476 neurons, 445 oligodendrocytes, 532 others) and there were 1628 annotated cells (187 astrocytes, 200 neurons, 783 oligodendrocytes, 458 others) in the basal ganglia. Sample size checks were carried out to ensure sufficient data to train the algorithms.

### 3.3.4. Held-out test set

Two slides, each from control and PSP groups, were randomly selected from frontal, temporal, parietal, occipital regions, and basal ganglia as held-out test slides. Non-occipital cortical slides were independently annotated by another neuropathologist (MB). F1-score was calculated with annotations from rater 1 as ground truth and supplemented with Cohen’s kappa for inter-rater agreement. For cortical regions, rater 1 has annotated 802 cells (101 astrocytes, 169 neurons, 231 oligodendrocytes, 301 others). Rater 2 has annotated 681 cells (162 astrocytes, 141 neurons, 229 oligodendrocytes, 149 others). For occipital region, 393 cells were annotated (62 astrocytes, 82 neurons, 98 oligodendrocytes, 151 others) and there were 417 annotated cells (56 astrocytes, 53 neurons, 155 oligodendrocytes, 153 others) in the basal ganglia.

**a) Region-specific tau classifier**




**b) Stratified 10-fold cross validation**

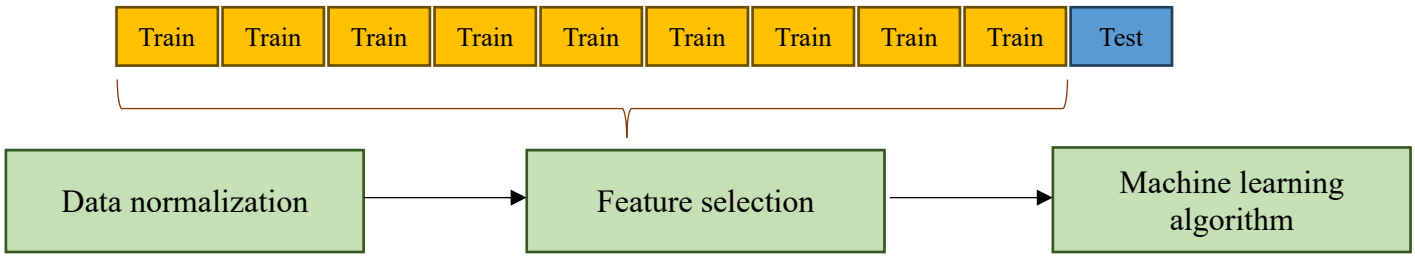



Figure 3.2: Schematic diagram showing a) annotated data and hyper-parameter tuning step for region-specific cell classifiers with resulting class-specific thresholds. b) In each loop through the stratified 10-fold cross validation, the following steps were carried out: data normalization, hyper-parameter tuning of feature selection approach and machine learning algorithm. Mean classification performance across 10 folds were used to evaluate pipeline performance and a set of parameters that yielded the highest performance across 10 folds was selected to further perform class-specific threshold-tuning.

### 3.3.5. Model development

First, for each training slide, mean haematoxylin intensity of the nucleus of all cells were normalized and cells with top 1% staining intensity were discarded to reduce DAB artefacts created from the bleeding of digital stain from hematoxylin to DAB channel. Annotated objects from all slides of the same brain regional grouping were pooled together, yielding a dataset. The dataset was then standardised (*mean = 0, SD =1*) and 10-fold stratified cross validation was used to train the region-specific cell classifiers. During the training phase, hyper-parameters of feature selection approaches (RFE, PCA) and machine learning algorithms (L-SVM, RBF-SVM, RF) were tuned (see Core methods). The machine learning algorithms were optimised based on the mean area under the 4 precision-recall curves (PR-AUC) using a one-vs-rest approach (Astro vs rest, Neuron vs rest, Oligo vs rest, Others vs rest).

Using the hyper-parameters found, optimal class-specific thresholds were tuned to tackle class imbalance for each brain regional grouping (non-occipital cortical regions, occipital region, and basal ganglia).

After classification, the precision, recall, macro F1-score and confusion matrix of the model were assessed. The model was then applied to the held-out test set to evaluate its performance and generalisability. Finally, the optimised model was applied to the remaining novel slides to perform tau classification and quantification for further analyses.

### 3.3.6. Cell quantification, tau positive and negative cells

Neuronal, glial, and other classes were quantified where key cells included neuronal and glial cells. The others class was mostly noise and would not be further analyzed. GNR and cell densities were calculated for all cells. GNR was defined as the ratio between glial cells (astrocytes and oligodendrocytes) and neurons. Cell density was defined for each cell type as the cell count divided by the area sampled ($\mu m^2$). Tau positive cells were those with DAB intensity equal to or greater than the tau positive threshold, otherwise they were tau negative cells (threshold = 0.27 in the mean DAB in the nucleus). The tau positive threshold was determined by considering the potential digital bleeding from hematoxylin channel to DAB channel as follows; pooling together slides across scanning batches, the top 10% most pigmented nuclei from hematoxylin staining intensity were selected from each control slide.

The 75$^{th}$ percentile value of DAB staining intensity in the nuclei from each slide was extracted and the maximum value across all control slides was inspected. The chosen threshold value was cross-checked with the most lightly stained heavy tau burden slide to ensure the validity of the chosen value.

### 3.3.7. Validation on novel slides

Boxplots using the *seaborn* package (Waskom, 2021) in Python were used to plot (1) GNR and regional cell type-specific density across brain regions for all subjects and (2) separately for control and PSP participants for comparison, (3) neuronal density of PSP participants across PSP pathological stage, (4) tau positive cell type-specific density of PSP participants, and (5) tau negative neuronal and astroglial density of PSP participants across PSP pathological stage. Further slide inspection was carried out and presented to explain the observed results.

## 3.4. Results

### 3.4.1. Comparison of feature selection methods using L-SVM

Table 3.6: Comparison of using no feature selection, using recursive feature elimination (RFE) wrapped around support vector machine with linear kernel (L-SVM) and principal component analysis (PCA) approaches. Selected features were subsequently fed to L-SVM classifier which was tuned for area under the precision-recall curve (PR-AUC) for cell classification. Precision, recall, F1-score without threshold-moving approach applied, and mean values from cross-validation with standard deviation in brackets are reported.

| No. features | Region | Precision | Recall | F1-score | PR-AUC |
|---|---|---|---|---|---|
| No feature selection | | | | | |
| 44 | Cortical | 0.76 (+/-0.04) | 0.81 (+/-0.04) | 0.76 (+/-0.04) | 0.82 (+/- 0.04) |
| 44 | Occipital | 0.77 (+/- 0.04) | 0.77 (+/- 0.04) | 0.76 (+/- 0.04) | 0.82 (+/- 0.05) |
| 44 | Basal ganglia | 0.75 (+/- 0.06) | 0.77 (+/- 0.06) | 0.75 (+/- 0.06) | 0.82 (+/- 0.06) |
| RFE | | | | | |
| 40 | Cortical | 0.76 (+/- 0.04) | 0.81 (+/- 0.03) | 0.76 (+/- 0.04) | 0.82 (+/- 0.04) |
| 30 | Occipital | 0.76 (+/- 0.04) | 0.77 (+/- 0.04) | 0.76 (+/- 0.05) | 0.83 (+/- 0.05) |
| 28 | Basal ganglia | 0.75 (+/- 0.06) | 0.78 (+/- 0.06) | 0.76 (+/- 0.06) | 0.82 (+/- 0.06) |
| PCA | | | | | |
| 24 | Cortical | 0.75 (+/- 0.05) | 0.80 (+/- 0.04) | 0.75 (+/-0.05) | 0.82 (+/- 0.04) |
| 24 | Occipital | 0.77 (+/- 0.05) | 0.77 (+/- 0.05) | 0.76 (+/- 0.05) | 0.83 (+/- 0.05) |
| 21 | Basal ganglia | 0.76 (+/- 0.06) | 0.78 (+/- 0.06) | 0.76 (+/- 0.07) | 0.82 (+/- 0.06) |

First, classification performances from L-SVM when full feature set, RFE wrapped around L-SVM, or PCA as the feature selection approach were compared across cortical, occipital, and basal ganglia (Table 3.6). The PCA approach selected at least 0.98 variance explained across all regions and retained smaller number of features or components than the RFE approach. The RFE approach retained the highest number of features for cell classification in cortical regions, followed by the occipital and basal ganglia regions. Classification performances with

or without feature selection approaches achieved the same PR-AUC across all regions. Therefore, RFE was chosen as the feature selection approach as PCA has issues with feature interpretability and using full feature set would mean retaining redundant features.

### 3.4.2. Comparison of machine learning algorithms

Table 3.7: Classification performance without the threshold-moving method from training for the cell classifiers (L-SVM, RBF-SVM, RF) with RFE as feature selection approach for the cortex, occipital region, and basal ganglia. Classifiers were tuned for area under the precision-recall curve (PR-AUC), where precision, recall and F1-score are also reported. Mean values from cross-validation are reported, and standard deviation in brackets.

| Region | Precision | Recall | F1-score | PR-AUC |
|---|---|---|---|---|
| L-SVM | | | | |
| Cortical | 0.76 (+/- 0.04) | 0.81 (+/- 0.03) | 0.76 (+/- 0.04) | 0.82 (+/- 0.04) |
| Occipital | 0.76 (+/- 0.04) | 0.77 (+/- 0.04) | 0.76 (+/- 0.05) | 0.83 (+/- 0.05) |
| Basal ganglia | 0.75 (+/- 0.06) | 0.78 (+/- 0.06) | 0.76 (+/- 0.06) | 0.82 (+/- 0.06) |
| RBF-SVM | | | | |
| Cortical | 0.75 (+/- 0.04) | 0.80 (+/- 0.04) | 0.76 (+/- 0.05) | 0.83 (+/- 0.04) |
| Occipital | 0.75 (+/- 0.04) | 0.77 (+/- 0.05) | 0.75 (+/- 0.04) | 0.83 (+/- 0.05) |
| Basal ganglia | 0.75 (+/- 0.06) | 0.77 (+/- 0.07) | 0.75 (+/- 0.07) | 0.85 (+/- 0.06) |
| RF | | | | |
| Cortical | 0.73 (+/-0.05) | 0.76 (+/-0.07) | 0.73 (+/-0.06) | 0.80 (+/-0.06) |
| Occipital | 0.75 (+/-0.04) | 0.76 (+/-0.04) | 0.74 (+/-0.04) | 0.82 (+/-0.05) |
| Basal ganglia | 0.76 (+/-0.05) | 0.77 (+/-0.06) | 0.75 (+/-0.06) | 0.85 (+/-0.06) |

Table 3.8: Classification performance with the threshold-moving method from training for the cell classifiers for the cortex, occipital region, and basal ganglia. Precision, recall and F1-score are reported. Mean values from cross-validation are reported and standard deviation in brackets.

| Region | Precision | Recall | F1-score |
|---|---|---|---|
| L-SVM | | | |
| Cortical | 0.78 (+/- 0.04) | 0.78 (+/- 0.06) | 0.77 (+/- 0.05) |
| Occipital | 0.79 (+/- 0.05) | 0.79 (+/- 0.05) | 0.79 (+/- 0.05) |
| Basal ganglia | 0.79 (+/- 0.05) | 0.78 (+/- 0.06) | 0.78 (+/-0.06) |
| RBF-SVM | | | |
| Cortical | 0.77 (+/- 0.04) | 0.78 (+/- 0.05) | 0.77 (+/- 0.04) |
| Occipital | 0.80 (+/- 0.07) | 0.79 (+/- 0.06) | 0.79 (+/- 0.06) |
| Basal ganglia | 0.79 (+/- 0.05) | 0.78 (+/- 0.07) | 0.78 (+/- 0.06) |
| RF | | | |
| Cortical | 0.78 (+/-0.05) | 0.80 (+/-0.06) | 0.78 (+/-0.06) |
| Occipital | 0.78 (+/-0.05) | 0.79 (+/-0.05) | 0.77 (+/-0.05) |
| Basal ganglia | 0.81 (+/-0.05) | 0.81 (+/-0.05) | 0.80 (+/-0.05) |

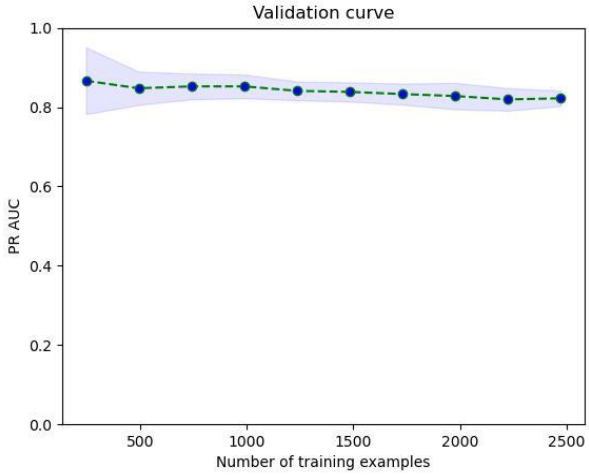
Next, using RFE approach as the feature selection method, the performance of L-SVM, RBF-SVM and RF were compared (Table *3.7*). PR-AUC across the three classifiers were very similar across all regions with at most 0.03 difference between mean PR-AUC values, which was still within the standard deviation around the mean. Threshold-moving approach was then applied to the three classifiers (Table 3.8). F1-scores across the three classifiers in all regions were very similar, with differences at most 0.02 where RF showed generally higher mean across regions than L-SVM and RBF-SVM. Therefore, RF was chosen as the algorithm to use forward as it is more interpretable and adaptable than SVM algorithms.

### 3.4.3. Classification performance of the chosen model

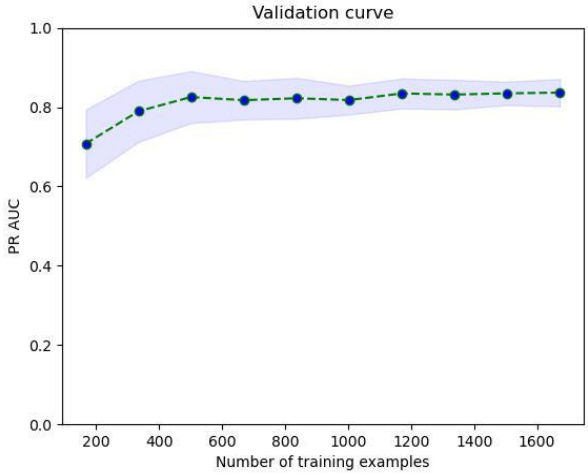
### 3.4.4. Sample size check

To ensure we had sufficient samples to train the RF classifiers for optimal results, I plotted the validation scores to check the stability of optimised PR-AUC from sampling 10-100% of the dataset incrementally (Figure 3.3). Across the 3 classifiers, PR-AUC from validation sets stabilised after 30% of the dataset was used, and the 95% confidence interval stayed small.

a) Cortical cell classifier




b) Occipital cell classifier




c) Basal ganglia cell classifier

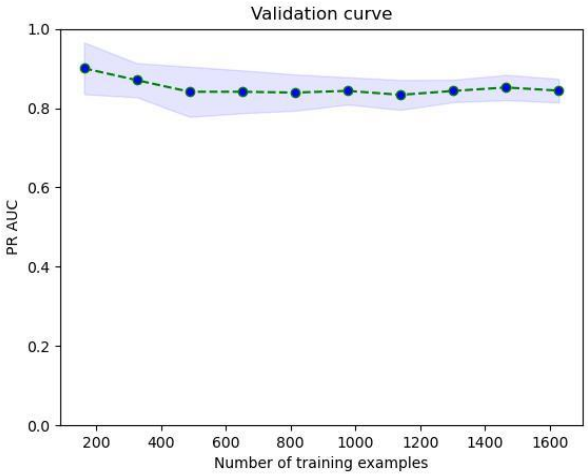



Figure 3.3: Plots showing mean PR-AUC score for each classifier from 10-fold stratified cross validation when 10-100% of the dataset was used. This shows the stability of the PR-AUC to inform whether increasing dataset would be useful in training the classifiers. Mean (dot) and 95% confidence intervals (shaded area) are shown.

### 3.4.5. Hyper-parameter tuning

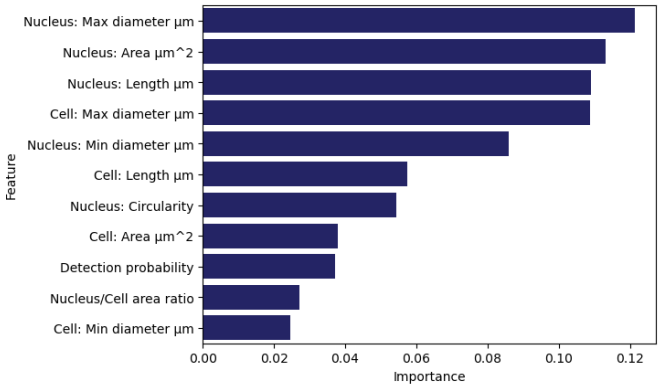
Table 3.9: Hyperparameter values from stratified 10-fold cross validation for the cell classifiers, specific to each regional grouping, including cortical regions (does not include occipital region), occipital region only and basal ganglia.

| Parameter | Cortical | Occipital | Basal ganglia |
|---|---|---|---|
| N_features_to_select | 38 | 28 | 38 |
| Sampling strategy | 'not majority' | 'auto' | 'not majority' |
| n_estimator | 900 | 200 | 600 |
| min_sample_split | 5 | 10 | 5 |
| min_sample_leaf | 4 | 4 | 2 |
| max_features | 0.2 | 0.4 | 0.2 |
| max_depth | None | 15 | 10 |
| max_sample | 0.5 | 0.25 | 0.75 |

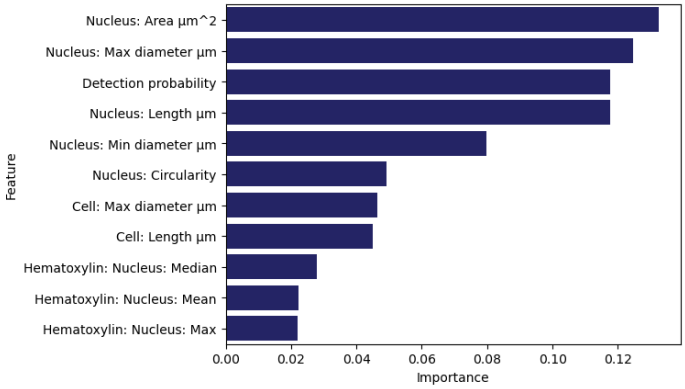
All classifiers were optimised for PR-AUC where cell classifiers for different regions yielded a different set of optimal hyperparameters (Table *3.9*). Cell classifiers in the cortex and basal ganglia selected the same number of features for cell classification (38 features), more than the occipital cell classifier which selected 28 features. All classifiers achieved PR-AUC scores above 0.80, where the basal ganglia cell classifier achieved the highest PR-AUC of 0.85, followed by occipital cell classifier of 0.82 and cortical cell classifier of 0.80 (Table *3.7*).

From inspecting the feature importance plots (Figure 3.4), nuclear morphology such as area and diameter are the top two most important features in distinguishing between different cell types. In general, morphological information were more important than haematoxylin staining intensities in classifying neuronal, glial, and other cell types across brain regions.

a) Cortical cell classifier




b) Occipital cell classifier




c) Basal ganglia cell classifier

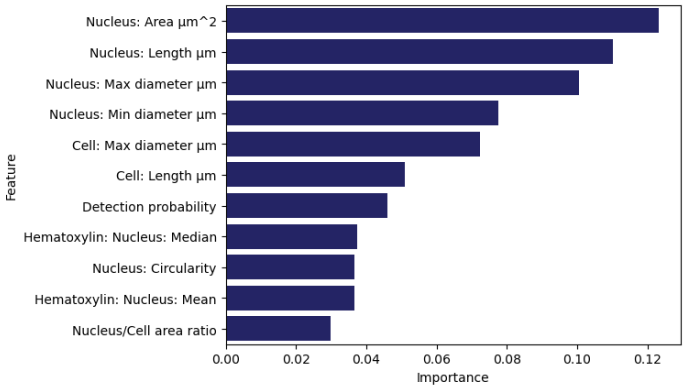



Figure 3.4: Top 10 most important features of each classifier from hyper-parameter tuning, a) cortical cell classifier, b) occipital cell classifier, c) cell classifier for basal ganglia.

### 3.4.6. Threshold-moving optimisation

Table 3.10: Class-specific thresholds for cell classifier for the cortex, occipital, and basal ganglia. Thresholds were optimised for F1-score using a one-vs-rest approach.

| Classifier | Astro | F1-score | Neuron | F1-score | Oligo | F1-score | Others | F1-score |
|---|---|---|---|---|---|---|---|---|
| Cortex | 0.51 | 0.58 | 0.51 | 0.91 | 0.28 | 0.86 | 0.28 | 0.77 |
| Occipital | 0.47 | 0.65 | 0.36 | 0.84 | 0.39 | 0.84 | 0.33 | 0.83 |
| Basal ganglia | 0.46 | 0.68 | 0.48 | 0.85 | 0.51 | 0.91 | 0.21 | 0.83 |

After optimising the hyper-parameters for each RF classifier, the next step is to tune class-specific thresholds for assigning labels to individual cells. Using a one-vs-rest approach, the class threshold with the highest F1-score was selected (Table *3.10*). The astrocyte and neuron classes had the same threshold of 0.51, and oligodendrocyte and others had the same threshold of 0.28 in the cortex. In the occipital region, the astrocyte class has the highest threshold (0.47), followed by oligodendrocyte (0.39), neurons (0.36), and others (0.33). For the basal ganglia, the oligodendrocyte class (0.51) had the highest class-specific threshold, followed by neurons (0.48), astrocytes (0.46), and others (0.21). In general, the astrocyte class had the smallest F1-score across the regions of below 0.7 while other cell types have class-specific thresholds with F1-score above 0.75. The proportion of cells being discarded as ambiguous was the highest in the cortex of 12% followed by the occipital region of 9% and basal ganglia of 7%. These results demonstrate the greater difficulty of classifying cell types in the cortex, as compared to the basal ganglia.

We compared the classification performance of the threshold-moving method to the default method of assigning class label based on maximum class scores (Table *3.7*, Table *3.8*). Applying threshold-moving method improved all metrics, specifically, F1-score by 0.03 for the occipital cell classifier and by 0.05 for the cell classifiers specific to the cortex and basal ganglia. Therefore, the threshold-moving methods were used for the final models.

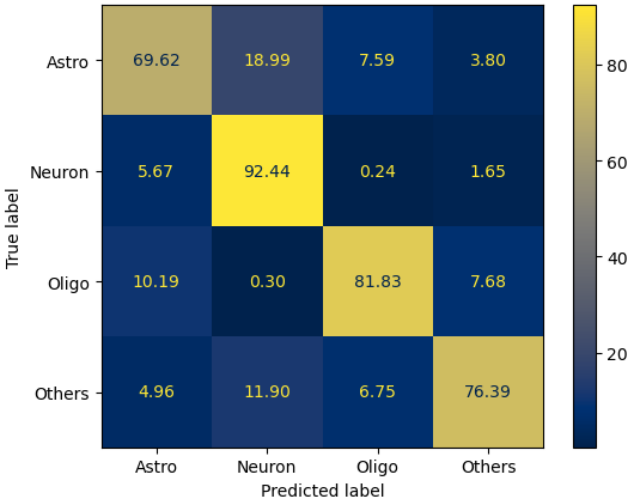
### 3.4.7. Confusion matrices

Confusion matrices for each classifier are shown in Figure 3.5. Cell classification in cortical regions achieved the highest accuracy of 92.44% in classifying neurons correctly, followed by oligodendrocytes (81.83%), others (76.39%) and astrocytes (69.62%). Neurons were most often misclassified as astrocytes (5.67%) while astrocytes were misclassified as neurons most often (18.99%), but also as oligodendrocytes (7.59%). Oligodendrocytes were most often misclassified as astrocytes (10.19%) and the misclassification of others was mostly as neurons (11.90%).

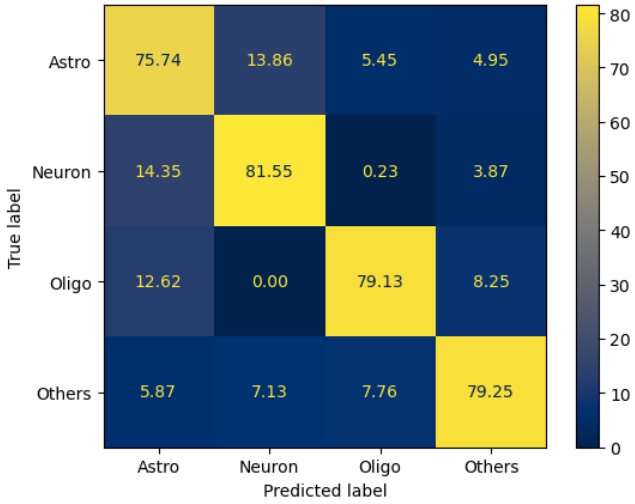
Similarly, cell classification in occipital cortex, the highest accuracy was achieved for the neuronal class (81.55%), followed by oligodendrocytes and others which yielded similar accuracies of 79.13% and 79.25% respectively. The accuracy for the astrocyte class was the lowest but higher than in cortical regions (75.74%). The mis-classification pattern followed the same pattern seen in cortical regions.

For the basal ganglia, the classifier performed best in the oligodendrocyte class (91.02%), followed by others (82.76%), neurons (81.77%) and astrocytes (68.00%). Oligodendrocytes here were most likely misclassified as others (5.63%) and others were also most likely misclassified as oligodendrocytes (8.13%). Neurons were misclassified as astrocytes most often (10.94%) and astrocytes were likely misclassified as neurons (14.86%) and others (12.00%).

a) Cortical cell classifier




b) Occipital cell classifier




c) Basal ganglia cell classifier

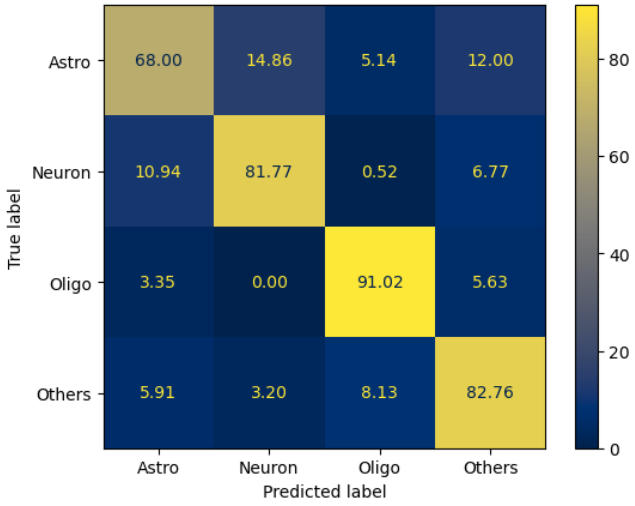



Figure 3.5: Mean confusion matrices from the 10-fold cross validation for a) cortical cell classifier, b) the occipital cell classifier, c) cell classifier for basal ganglia.

### 3.4.8. Validation on a held-out test set

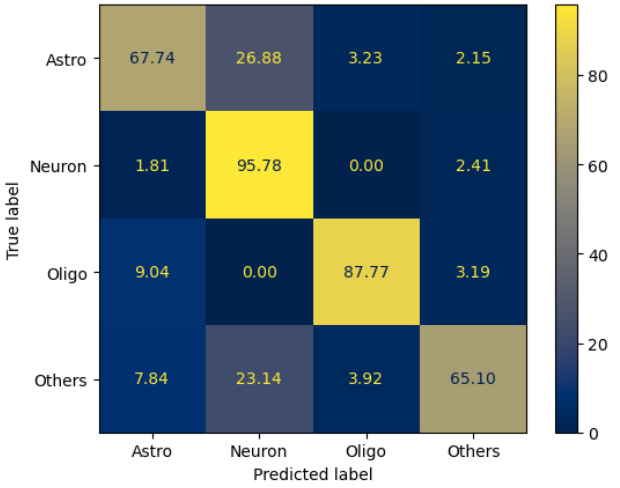
Table 3.11: Classification performance on a held-out test set from cortical, occipital regions and basal ganglia. Precision, recall and F1-score are reported.

| Region | Precision | Recall | F1-score |
|---|---|---|---|
| Cortex | 0.78 | 0.79 | 0.77 |
| Occipital | 0.79 | 0.80 | 0.79 |
| Basal ganglia | 0.81 | 0.85 | 0.83 |

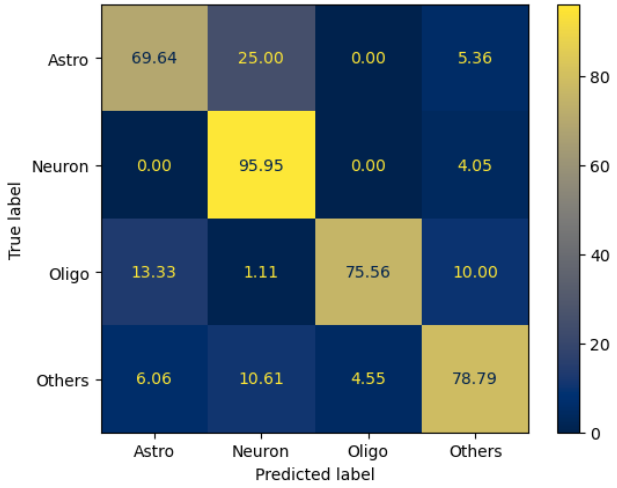
For the held-out test set, the cell classifier for the basal ganglia achieved the highest F1-score of 0.83, followed by the classifier for occipital cortex of 0.79 and cortical regions of 0.77 (Table *3.11*). For inter-rater agreement in cortical regions, F1-score was 0.85 and Cohen's kappa was 0.81.

From inspecting the confusion matrices (Figure 3.6), the cortical cell classifier achieved the highest accuracy for neurons (95.78%), followed by oligodendrocytes (87.77%), astrocytes (67.74%), and others (65.10%). For the occipital region, neuron class achieved the highest accuracy (95.95%), followed by others (78.79%), oligodendrocytes (75.56%) and astrocytes (69.64%). This order was the same for cell classification in basal ganglia, except the accuracy for oligodendrocyte class (89.44%) was higher than the others class (79.85%). As before, the astrocyte class achieved the lowest accuracy across the brain regions where the lowest was in the cortical regions. Examples of correct and incorrect classifications are shown in Figure 3.7 and Figure 3.8, illustrating heterogeneity within each cell class which can impact classification accuracy. These examples, along with feature importance plots (Figure 3.4), illustrate that the classifiers tend to rely more on nuclear size than other features for cell classification, which could help explain the misclassification pattern observed.

a) Non-occipital cortex




b) Occipital cortex




c) Basal ganglia

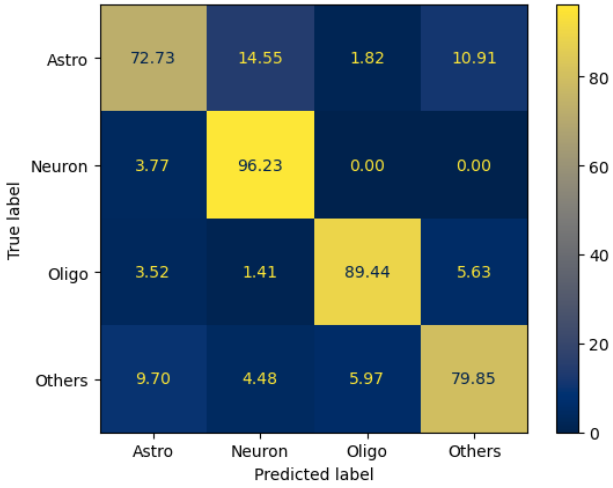
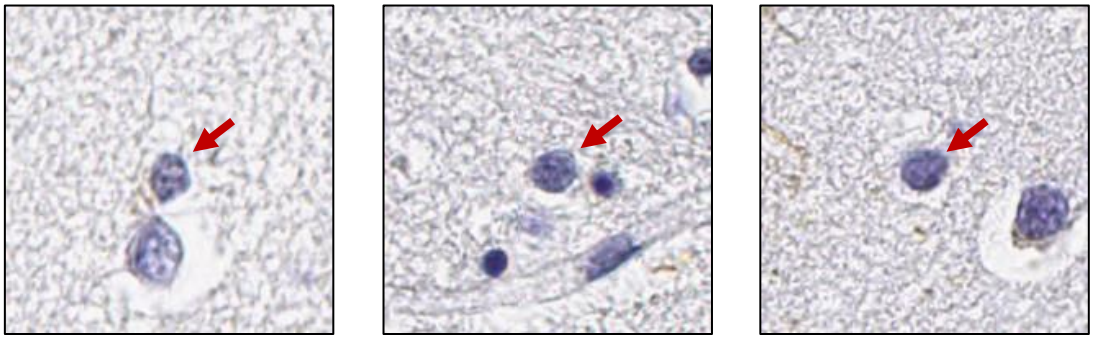
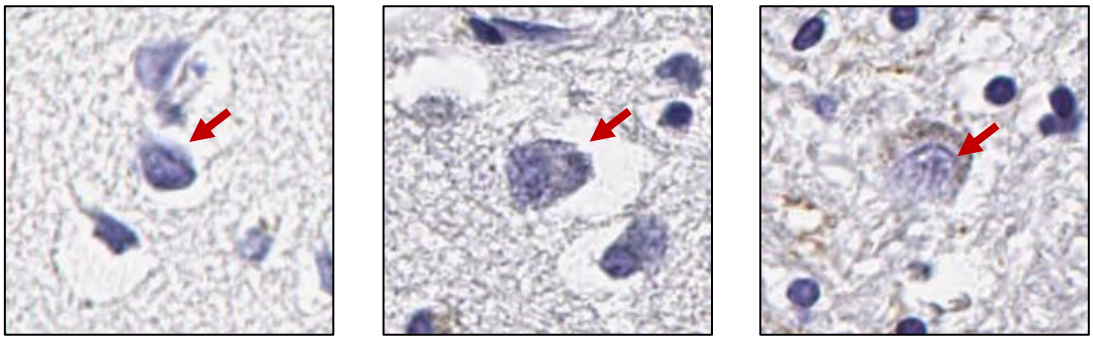



Figure 3.6: Confusion matrices of held-out test set by region a) non-occipital cortex, b) occipital cortex, c) basal ganglia.

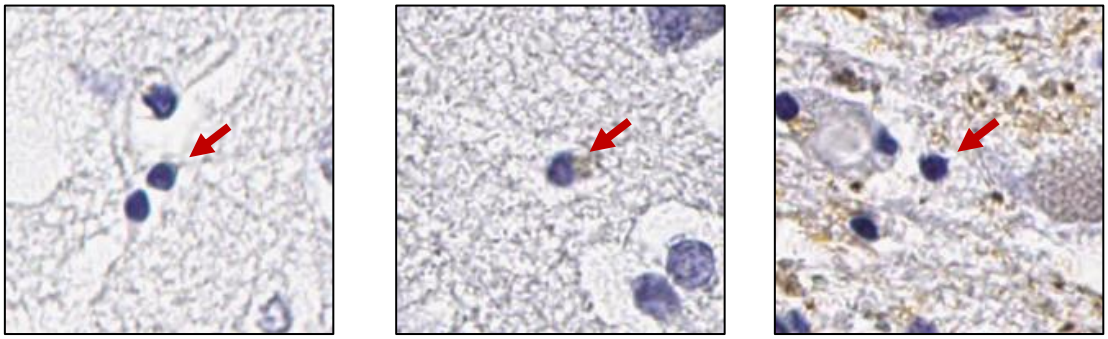
Correct astrocyte classification



Correct neuron classification



Correct oligodendrocyte classification



Correct others classification

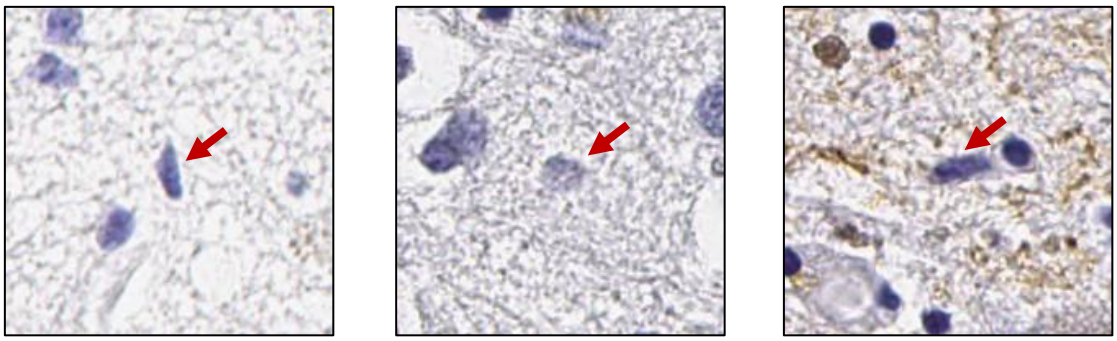


Figure 3.7: Examples of correct classification from held-out test sets for each cell type drawn from cortical regions (left column), occipital region (middle column), and basal ganglia (right column). Arrows point at cell of interest.

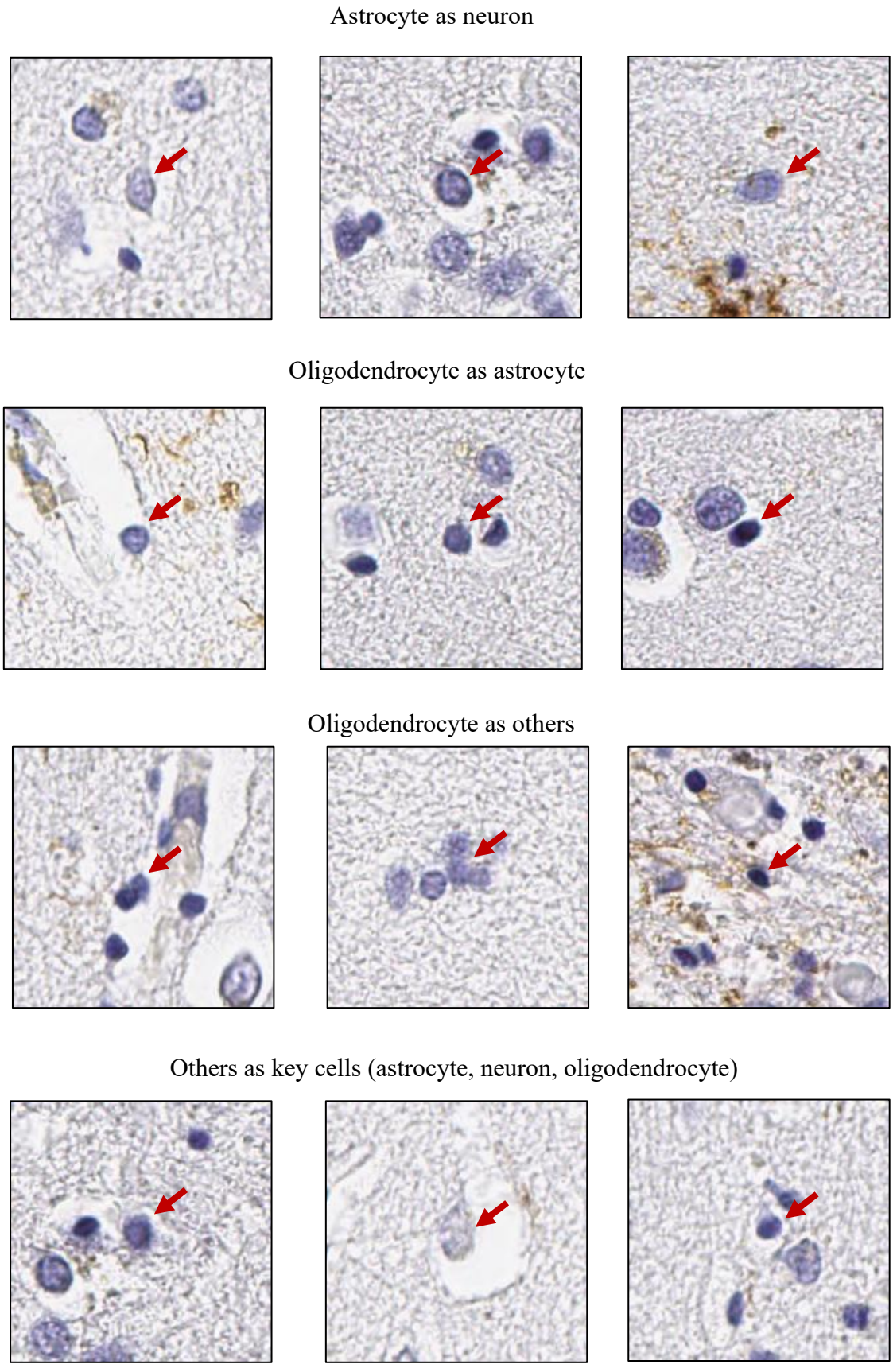



Figure 3.8: Main misclassification between cell types from held-out test sets (truth label as predicted label), drawn from cortical regions (left column), occipital region (middle column), and basal ganglia (right column). Except for Others as key cells that are only drawn from cortical regions as examples.

### 3.4.9. All subjects: total GNR and cell density

The resulting pipeline was applied to all novel slides for validation. GNR was calculated for all brain regions (Figure 3.9). Mean GNR in cortical regions ranged from 1.21 (parietal) to 1.56 (temporal) where GNR was much higher in the basal ganglia and the highest GNR was found in the globus pallidus (15.40), followed by subthalamic nucleus (10.62) and putamen (3.43). In general, others density (which include artefacts and endothelial cells) was the highest across all cell type-specific densities across all brain regions, except for putamen. When inspecting key cell density across cortical regions, neuronal density was the greatest as compared to glial density, followed by oligodendroglial and astrocytic density being the lowest. In contrast, oligodendroglial density was the highest in basal ganglia, followed by astrocytic and neuronal density for subthalamic nucleus and globus pallidus. In putamen, higher neuronal density than astrocytic density was observed.

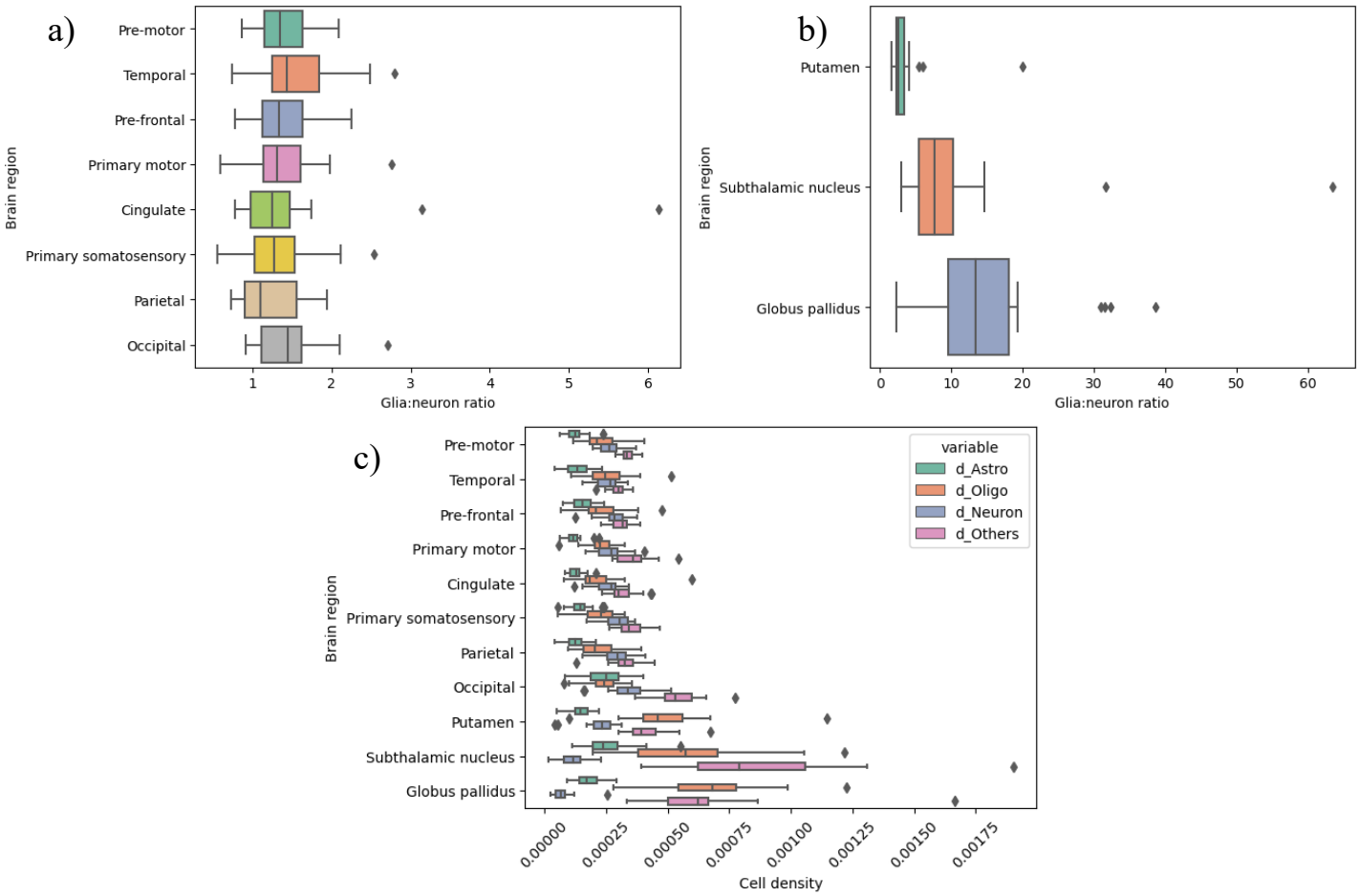



Figure 3.9: Boxplots showing a) glia:neuron ratio in cortical regions based on a random forest classifier and b) basal ganglia nuclei, and c) cell type-specific density for each brain region. d_Astro astrocytic density, d_Oligo oligodendroglial density, d_Neuron neuronal density, d_Others others density.

### 3.4.10. Control vs PSP: total key cell type-specific density and GNR

Cell type-specific density and GNR were compared between PSP and controls (Figure 3.10), it is difficult to interpret results for the basal ganglia regions as only 1 control has those regions. Focusing on cortical regions, controls showed higher mean key cell density (all cell types excluding others class) in temporal, parietal and occipital regions than PSP subjects, while PSP showed higher key cell density in the pre-frontal region. Furthermore, PSP showed higher mean astrocytic and neuronal density across cortical regions than controls, except for the occipital region where the opposite pattern is observed. PSP tended to show lower mean oligodendroglial density than controls across cortical regions. For GNR across cortical regions, PSP showed lower GNR in temporal, parietal and occipital regions, similar GNR in pre-frontal and higher GNR in cingulate than controls.

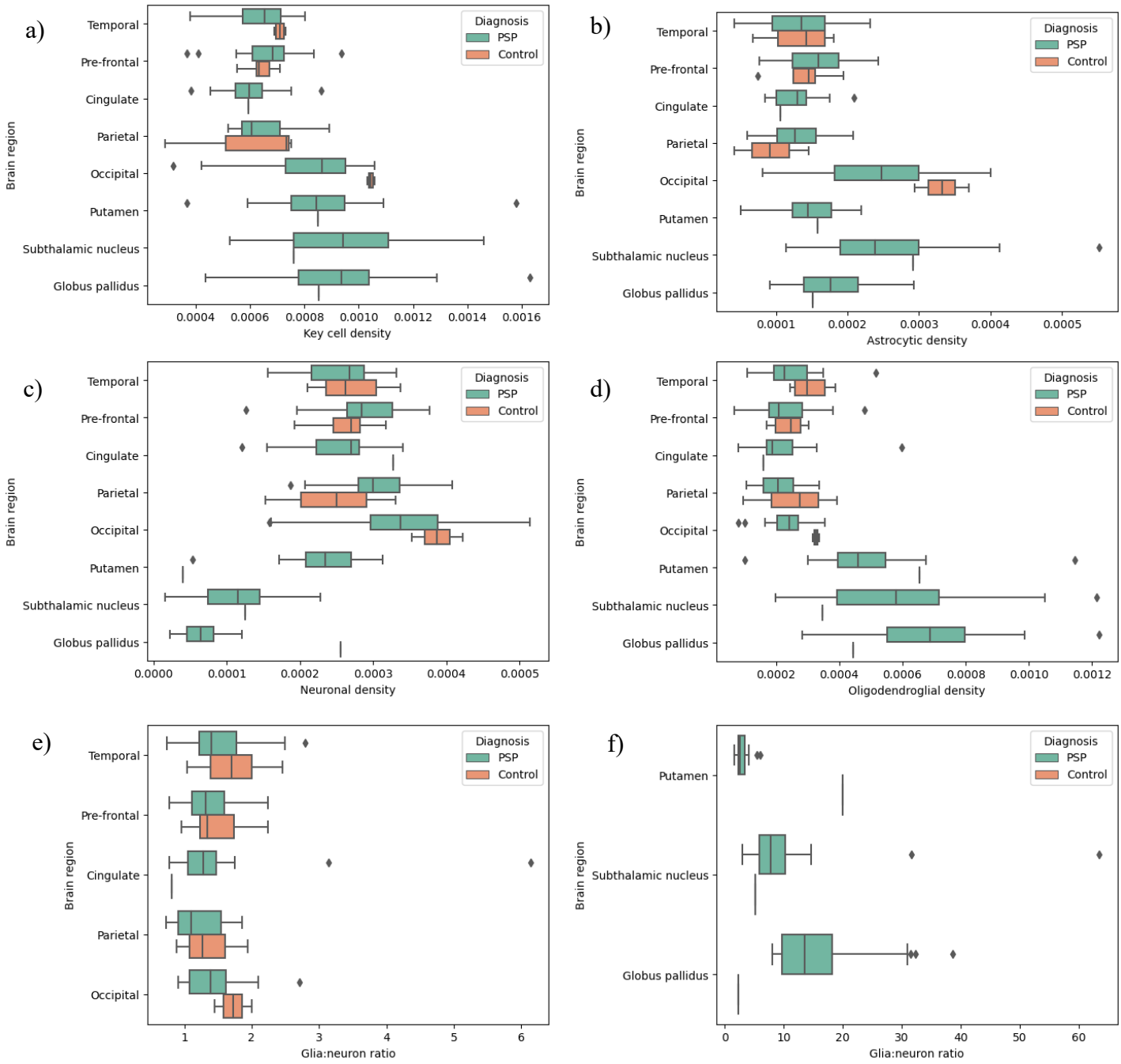



Figure 3.10: Boxplots showing a) key cell type-specific density (neuronal and glial cells), b) astrocytic density, c) neuronal density, d) oligodendroglial density each brain region, e) GNR in cortical regions, f) GNR in subcortical regions between PSP and control participants.

### 3.4.11. Total neuronal density in PSP, by PSP stage

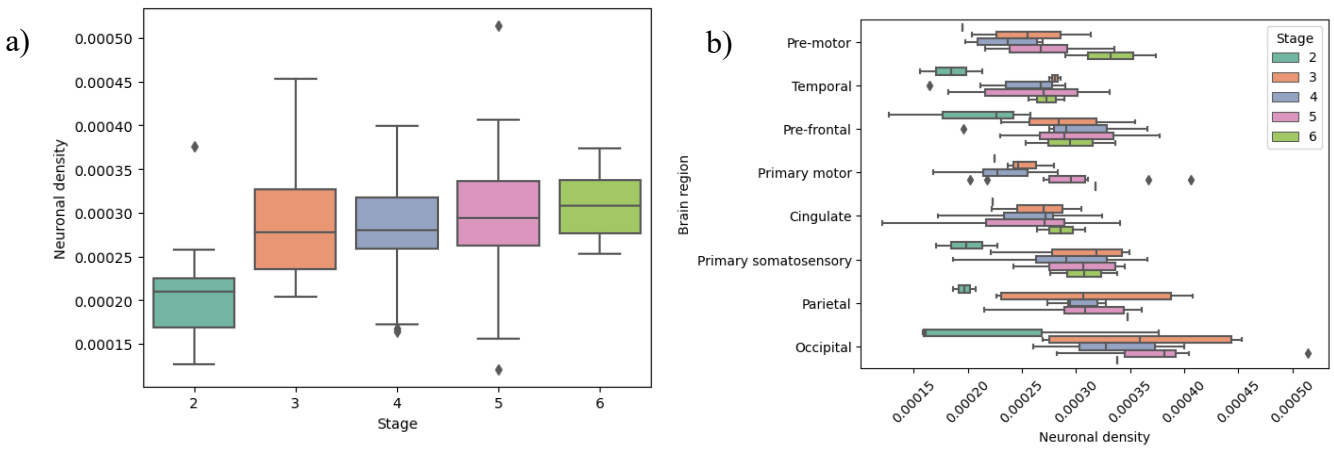



Figure 3.11: Boxplots showing a) total neuronal density for each PSP pathological stage from 2-6 and further b) separated by region.

Neuronal density across PSP stage was assessed to investigate for neuronal loss where stage 2 showed the lowest neuronal density as compared to later PSP stages which showed similar but a trend towards increasing in neuronal density from stage 3-6 (Figure 3.11). Looking more closely at each region from subjects with available PSP stage, the neuronal density in stage 2 was generally the lowest, and followed the pattern observed from stages 3-6 across brain regions.

### 3.4.12. Robustness towards tau aggregation: tau positive cells

First, tau positive cell density in PSP was plotted across cortical and basal ganglia regions (Figure 3.12). Basal ganglia nuclei showed higher tau positive cell density than cortical regions, specifically tau positive oligodendroglial density. In cortical regions, tau positive oligodendroglial density was the highest, and tau positive neuronal and astrocytic density were relatively similar.

I further inspected PSP slides with high tau density to test the robustness of the classifier for cell types with tau aggregates (Figure *3.13*). In general, nuclei detection was influenced by tau aggregation, resulting in multiple smaller fragmented nuclei detections, or a single enlarged detection. Tufted astrocytes were often detected as neurons or multiple others when

nuclei detection failed to detect the cell body of astrocytes. For neurofibrillary tangles, they were still largely detected as a single neuron, but could also be detected as multiple cells and labelled as a neuron or others. Furthermore, coiled bodies might be detected as a neuron or others, depending on whether the nuclear detection was enlarged by the tau inclusion or not. Lastly, neuronal threads that were not associated with a cell body were either detected from the nuclei detection step and classified as others or were undetected. These findings suggests that the automated cell classification is unreliable in tau containing cells.

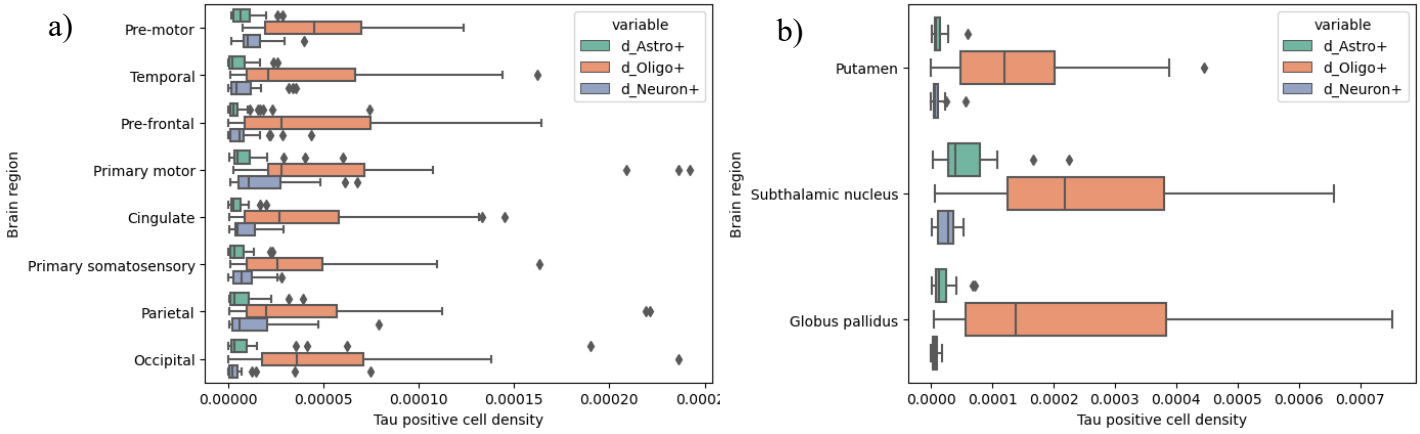



Figure 3.12: Boxplots showing tau positive cell type-specific density in PSP participants across a) cortical regions, and b) basal ganglia nuclei. d_Astro+ tau positive astrocytic density, d_Oligo+ tau positive oligodendroglial density, d_Neuron+ tau positive neuronal density.

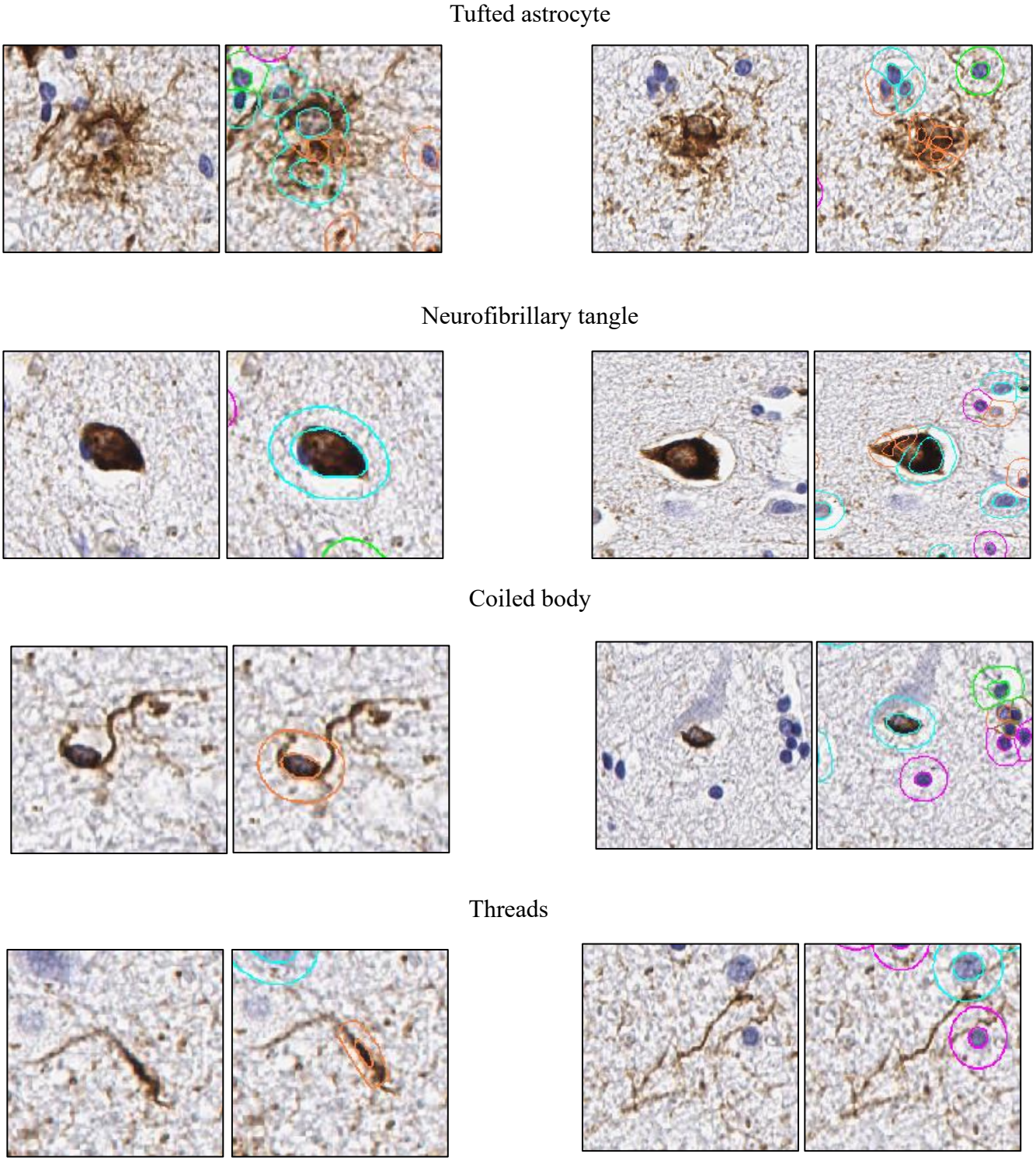



Figure 3.13: Examples of correct and incorrect cell classification result overlayed on tau pathological hallmarks (tufted astrocyte, neurofibrillary tangle, coiled body, and threads) in PSP. Prediction of the cell type is colour coded: neuron (Cyan), oligodendrocyte (magenta), astrocyte (green), others (orange).

### 3.4.13. Tau positive threshold selection

Slides from all batches available were pooled together and the tau positive threshold was determined from control slides with heavy haematoxylin staining and cross-checked with PSP slides with light tau aggregation (threshold=0.27). We further inspected control and PSP slides with the lowest and highest staining intensity to ensure the threshold value generalised at both extremes. Based on Figure 3.14, tau negative cells were largely captured with the chosen threshold as only tau aggregates were detected as tau positive. Nevertheless, on occasion where the bleeding between DAB and haematoxylin channels was severe, highly pigmented healthy cells in the control slide could be detected as tau positive. This issue was likely tackled by the initial discarding of cells where cells with top 1% haematoxylin staining intensity per slide were discarded before cell classification step. This was rarely the case with PSP slides that generally had lower staining intensity across cells than control slides.

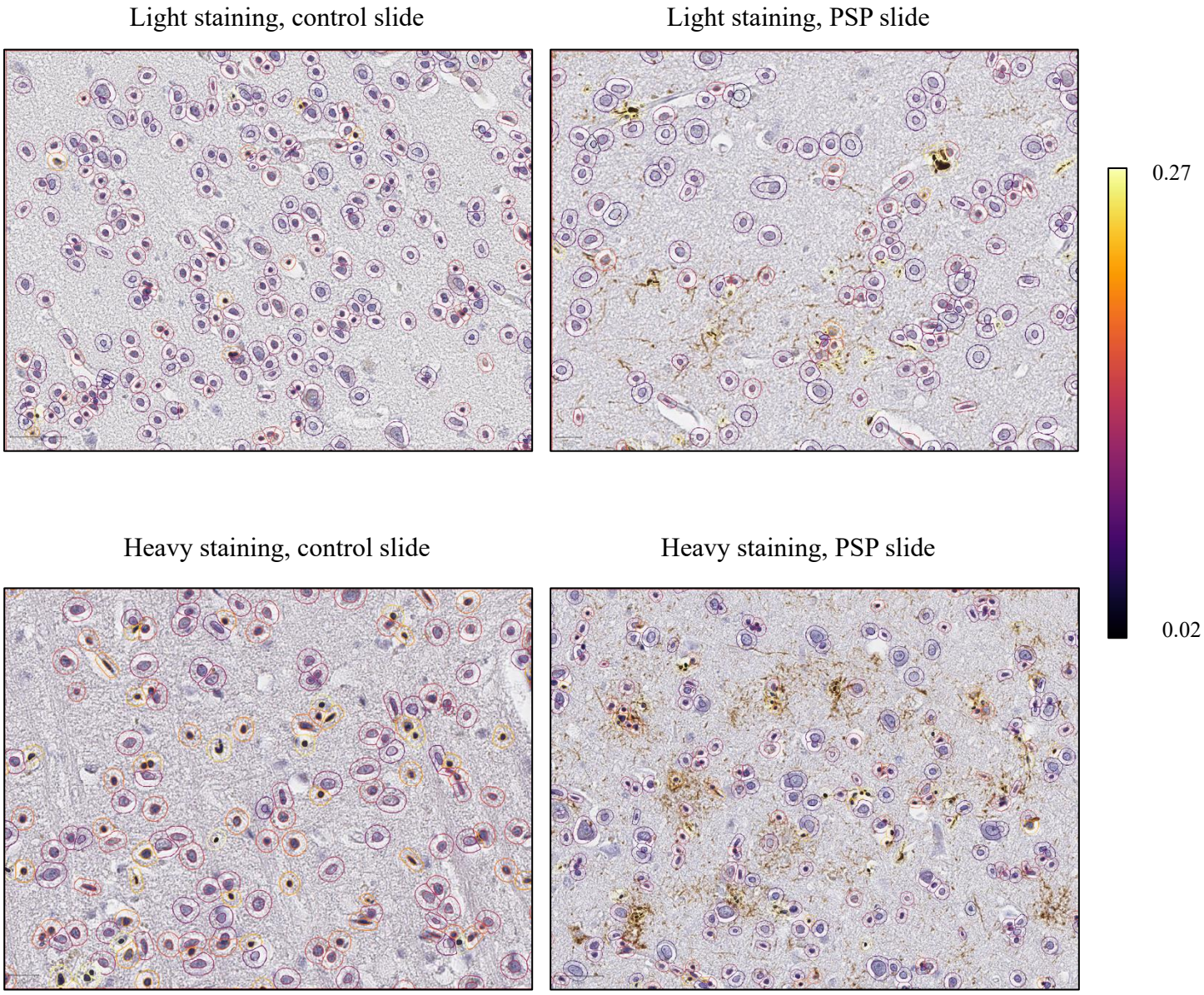



Figure 3.14: With tau positive threshold of 0.27, slides with the lowest and highest haematoxylin staining intensity from control and PSP were inspected to ensure the robustness of the threshold value. Cells labelled in black are tau negative, and in pale yellow are tau positive. The colour bar represents the intensity of DAB staining in the nucleus.

### 3.4.14. Robustness in out-of-sample tau negative cells

We plotted tau negative neuronal and astrocytic density (Figure 3.15) to further inspect the unexpected finding seen in Figure 3.11 that neuronal density in stage 6 appeared higher than stage 2. We found that the same pattern was observed with tau negative neuronal density and astrocytic density from stage 2-6 appear relatively unchanged but towards a decreasing trend. We inspected a pre-motor slide from a PSP stage 6 participant (Figure 3.16) and found substantial misclassification of astrocytes as neurons, explaining this unexpected finding.

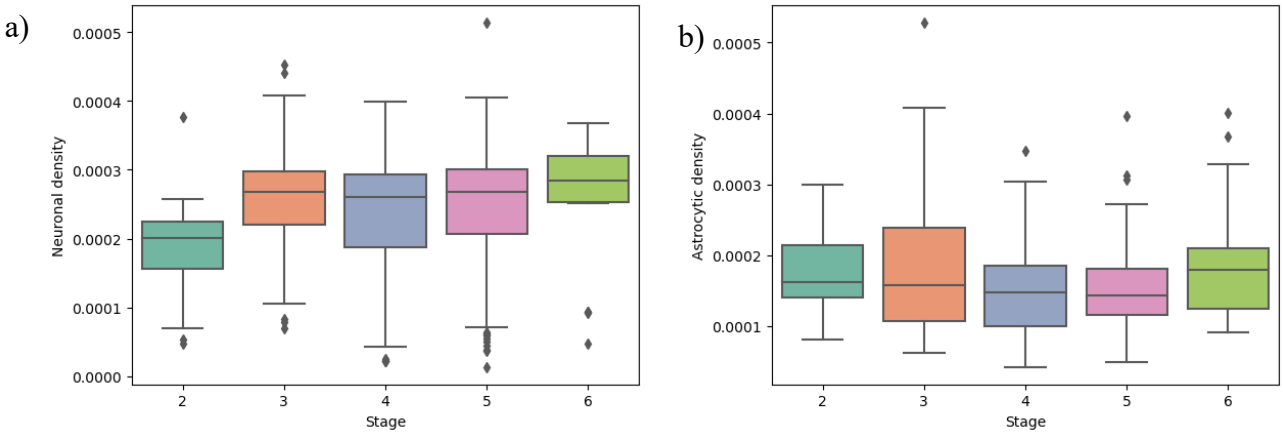



Figure 3.15: Boxplots showing tau negative a) neuronal density and b) astrocytic density for each PSP pathological stage from 2-6.

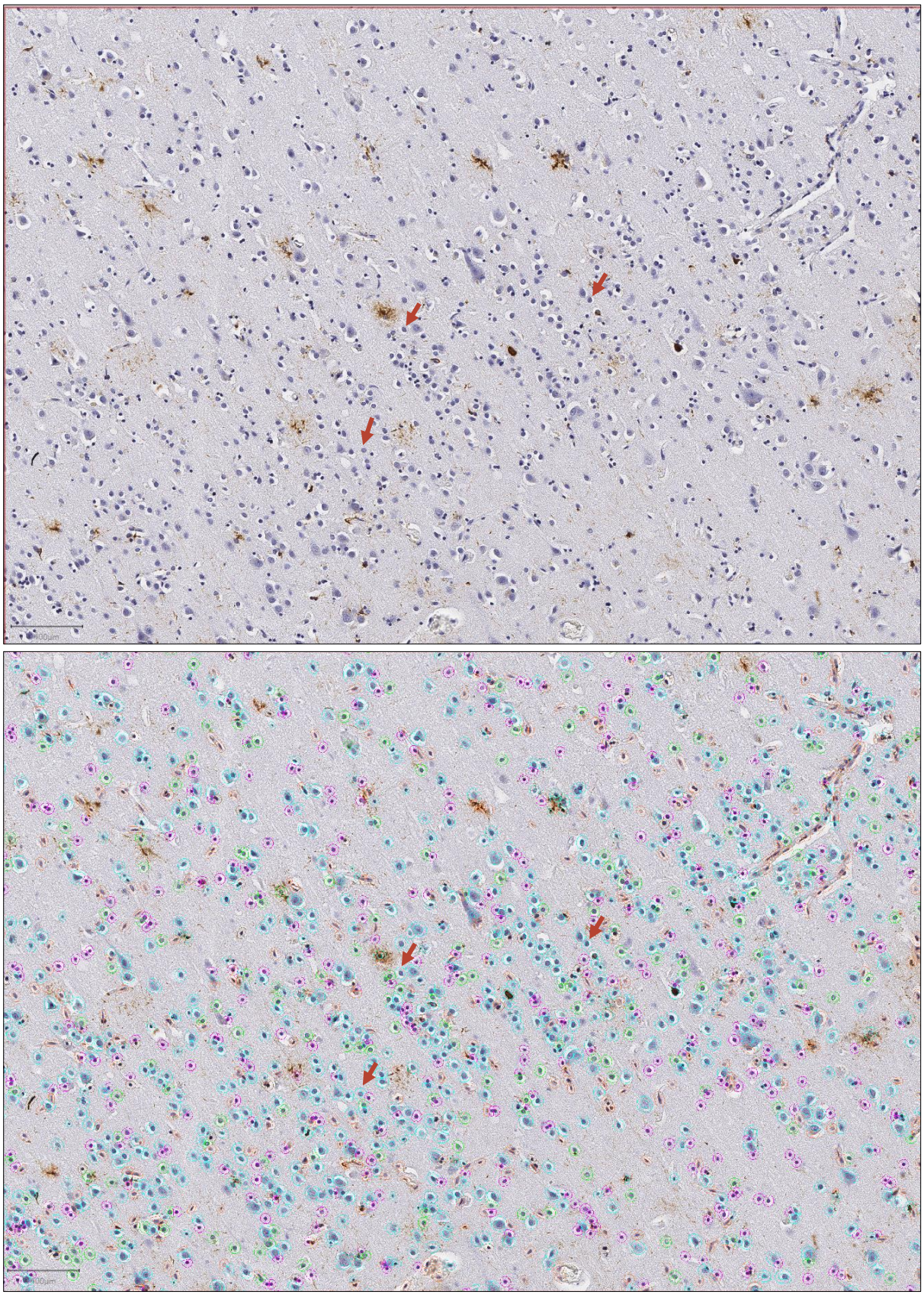


Figure 3.16: A pre-motor slide of PSP stage 6 donor participant, showing misclassification of astrocytes as neurons where examples of misclassifications are indicated by red arrow. Predicted as neurons (blue), astrocytes (green), oligodendrocytes (magenta), others (orange).

## 3.5. Discussion

### 3.5.1. Overview

I have attempted to develop a pipeline for total and tau positive cell type-specific density in *post mortem* cortical and basal ganglia slides in PSP and control donors.

First, I compared multiple feature selection approaches and machine learning algorithms to determine the most suitable configurations for the cell classification pipeline. Classification performances across approaches were very similar, therefore, the decision was made based on other criteria concerning pipeline interpretability and adaptability. I have shown that PCA retained the lowest number of features and classification performance of L-SVM was the same whether feature selection was applied or not. This suggests that even though using the full feature set does not add new information, it also does not introduce noise that impairs classification performance. PCA was the most effective approach to remove redundancy and retained important information for cell classification. Even though RFE retains more features than PCA, it enables the pipeline to be more interpretable as it does not involve feature decomposition and still reduces redundancy compared to using all the available features. It was therefore selected as the feature selection approach. I further compared classification performance across 3 machine learning algorithms: L-SVM, RBF-SVM and RF. They performed similarly with and without the threshold-moving approach applied. RF was selected as the algorithm for cell classification as it is more adaptable or flexible than L-SVM and is more interpretable than RBF-SVM in which feature importance cannot be natively extracted.

With RF with RFE as the chosen configuration, the classification performance differs across brain regions, where the highest PR-AUC was achieved in the basal ganglia, followed by occipital and cortical regions. I have shown that a threshold-moving method improves classification performance by adapting class-specific thresholds to take account of class imbalance within each brain region. Nevertheless, class-specific accuracy from the final models differed drastically, especially when comparing the majority (such as neurons) to the minority cell class (such as astrocytes) for each brain grouping where the former achieved

much higher accuracy than the latter. The results were further confirmed by the held-out test set and F1-score achieved by the pipeline (0.77) was lower than an independent expert neuropathologist (0.85) in cortical regions when compared against the primary neuropathologist.

We applied the pipeline to all available slides to gauge whether classification performance would be sufficient for further analyses. We found that GNR and cell type-specific density in the cortex were in line with existing literature (von Bartheld et al., 2016). Nevertheless, PSP participants did not consistently show higher GNR than controls across cortical regions as anticipated. Still, tau positive cell density was higher in basal ganglia nuclei than cortical regions in PSP. However, neuronal density appeared to be lowest in pathological PSP stage 2 compared to more severe PSP stages. Upon slide inspection, these results were explained by the misclassification of cells with tau aggregation which distorted the nuclear detection step. When nucleus detection was falsely enlarged, glial cells were detected as neurons, while when the nuclei detection step resulted in multiple detections of a single cell, they were generally classified as others due to their small and fragmented nature. Moreover, astrocytes were often misclassified as neurons; and higher stages of pathology have a great number of reactive astrocytes that are more likely to be misclassified as neurons. (Li et al., 2019). This may also contribute to the surprising increase in neuronal density with pathological stage. To be certain, we would need to additional astrocytic staining to objectively determine the misclassification rate and astrocytic density.

### 3.5.2. Challenges and solutions

There were 3 main challenges that I addressed in this study. First, there is a class ratio imbalance of neuronal and glial cells in the brain. For example, oligodendrocytes are the most abundant cell type in the cortex, followed by neurons and astrocytes (von Bartheld et al., 2016), (WallÃ¸e et al., 2014), as reflected by our annotated dataset. Class imbalance is a common issue of real-world datasets which most machine learning algorithms struggle with since they are designed to deal with balanced datasets (Brownlee, 2020).With imbalanced datasets, machine learning algorithms are often biased towards the majority class. This is a problem as the lower accuracy from the minority class will influence the classification performance of the model (Basha et al., 2022).

Second, the ratio of class imbalance and cell morphologies differ between brain regions. GNR in the prefrontal cerebral cortex is roughly 1.5 but can range from 1.2 (occipital) to 3.6 in frontal areas (Ribeiro et al., 2013; von Bartheld et al., 2016). Neurons in the occipital cortex are generally smaller in size and are more densely packed (Ribeiro et al., 2013). Furthermore, in the basal ganglia, astrocytes are more abundant than neurons (Salvesen et al., 2015), which is opposite to the cerebral cortex.

Third, classifying cells on bright-field images with high precision is difficult as the slide preparation technique does not fully capture the biological diversity of different cell types, resulting in ambiguity (Oei et al., 2019).

To tackle the class imbalance challenge, we chose a probabilistic classifier which would enable us to tune class-specific thresholds. We chose a relatively simple but powerful machine learning algorithm; the random forest which yielded highly satisfactory results for our task. Major advantages of using a random forest as opposed to more complex algorithms, such as a neural network, is that it requires less training data, tuning time, and is readily interpretable (Kong & Yu, 2018) . Not only that, random forest algorithms are effective when there are highly correlated predictors (an advantage over linear regression approaches) and can capture non-linear relationship between the predictors and outcome (Boulesteix et al., 2012). This is highly relevant to our task where features relating to cell morphology and staining intensities from various digital channels are likely highly correlated. With the supervised learning approach, we could carefully annotate a highly accurate dataset to train the classifier within a reasonable timeframe and extract feature importance which provides insights into how the algorithm works, providing greater confidence in the results. Moreover, since most machine learning algorithms struggle with learning from an imbalanced dataset, I chose to use a balanced random forest classifier to create a balanced bootstrap training dataset for the classifier (Chen, 2011; Luo et al., 2019; More & Rana, 2017). Instead of annotating the dataset in a class-balanced manner, I chose to systematically sample regions on the slide to annotate all tau objects that we are confident of, to reflect the actual class distribution. The threshold-moving method was then used to fine tune the classifier to handle class imbalance. This helped to ensure that our training dataset reflected the class distribution of the actual dataset, and that the classification performance would generalise to other datasets.

To tackle with the second challenge of region-specific class ratio and tau morphology, we grouped similar brain regions together and trained region-specific classifiers. This resulted in 3 region-specific cell classifiers for the cortical, occipital, and basal ganglia regions.

Lastly, to tackle with the ambiguity inherent in identifying cells, we employed an approach used by neuropathologists and introduced an 'ambiguous' class for when the classifier was unsure of its decision. An object was labelled as ambiguous when the class scores passed more than one class-specific threshold, or none. This captured both the upper level of ambiguity where the classifier thinks an object could belong to more than 1 class, and the lower end of ambiguity where the classifier thinks this object does not resemble any of the pre-defined classes. Therefore, the classifier does not need to force a label and risk making a mistake.

### 3.5.3. Outstanding challenges

Despite our attempts, nuclear detection step is influenced by tau aggregation so that classifying tau positive cells with high accuracy was not possible with the current pipeline. The current pipeline has the potential to work well with healthy neuronal and glial cells, but there are two main issues that we believe hinder this. First, the current feature choices may be inadequate for accurate cell classification. The astrocyte class consistently achieved the lowest classification accuracy across the cortex and basal ganglia. Despite it being the minority class, another potential explanation could be that the characteristics of astrocytes lie between neurons and oligodendrocytes. Based on feature importance plots, cellular morphology such as the nucleus area was one of the top 2 most important features in classifying cell types. Neurons are slightly bigger than astrocytes, while oligodendrocytes are the smallest and differ from astrocytes considerably. Misclassifications happened when large oligodendrocytes resembled small astrocytes, and large astrocytes resembled smaller neurons. As a result, astrocytes were most often misclassified as neurons and sometimes as oligodendrocytes while misclassification between oligodendrocytes and neurons was rare.

The confusion between key cell types and others could come from the disagreement between the algorithm and the pathologists as to whether a given cell should be ignored. Therefore, more informative, and distinctive features of each class would be useful to enable better classification. It may be useful to directly use pixel values of the cells and feed into more

complex machine learning algorithms such as convolution neural network to better classify between cell types.

Second, the currently available data may lack the information required to accurately classify cell types. The most challenging aspect of the pipeline development was accurately distinguishing between cell types by the human raters, whether that be cell annotation by the pathologists which was very time-consuming or cell type recognition by the data scientist. This points to the fact that slides stained with haematoxylin for nuclei detection may not readily provide sufficient information to distinguish between cell types reliably. Haematoxylin staining only enables cell nuclei visualisation, therefore other important cytological features such as cytoplasm, heterochromatin, and the nucleolus are missing. Such information could help, for example, to distinguish neurons from astrocytes since neurons would show a patent rim of cytoplasm encircling the nucleus, which would be possible with Nissl staining (Garcia et al., 2018). Therefore, cell type-specific staining agent may be required to further improve on the confidence of annotated dataset and for better understanding of the classification results by data scientists. Any choice of preparation technique should still aim for a simple solution that can be widely adopted and scalable.

Overall, we feel that the current pipeline is not ready to be taken forward to analyse tau positive cells. Therefore, we aimed to re-design the pipeline to focus on tau hallmarks in PSP, outlined in the next chapter.

# Chapter 4: Tau type-specific quantification

## 4.1. Introduction

The characterization of tau pathological hallmarks is essential in permitting the study of neurodegenerative tauopathies (Kovacs, 2015; Kovacs et al., 2020). Recent efforts in using automated pipelines to characterize tau burden in tauopathies can largely be grouped into 3 categories which are: 1) the quantification of total tau burden, 2) classifying multiple tau type-specific aggregates, and 3) classifying a single tau aggregate type. Studies focusing on total tau burden have quantified percentage area covered by tau without cell type-specific distinction in both grey and white matter to investigate disease-specific signature across tauopathies (Coughlin et al., 2022; Vega et al., 2021). Specifically, only in grey matter, studies have largely focused on tau hallmark classification to aid postmortem diagnosis. Studies have attempted to differentiate between AD and non-AD tauopathies such as PSP, CBD and PiD by quantifying disease-specific tau hallmarks (Koga et al., 2021, 2022). Amongst the tau aggregate types, NFT has been most widely researched, most likely because it is found across tauopathies, especially in AD which is the most common type of dementia (Y. Zhang et al., 2022). Studies have attempted to quantify and investigate NFT counts and cognitive impairment (Marx et al., 2022), as well as developing a novel technique that can generate 3-dimensional mapping of NFTs in the brain (Yushkevich et al., 2021).

However, these studies are largely at proof-of-concept stage where only a small subset of brain regions were investigated (Koga et al., 2021, 2022; Signaevsky et al., 2019). For investigating the tau distribution pattern, it is desirable to include a wide range of brain regions to enable comparison between regions. Studies that have developed a pipeline for tau type-specific quantification have also largely focused on key disease-specific pathological hallmarks that would be useful for distinguishing between different tauopathies (Koga et al., 2021, 2022). Tau aggregate types that are also informative but may not be directly useful for diagnosis were not quantified. There remains a gap in the literature for a digital pipeline that can cover a larger number of regions and more detailed quantification for mechanistic studies.

### 4.1.1. The present study

Here, I aimed to quantify tau pathology in PSP *post mortem* brains by developing a digital tau pathology pipeline for whole slide images. This pipeline (Figure 4.1) has been developed to work with brain regions included in the current consensus PSP pathology staging scheme (Kovacs et al., 2020) and additional cortical regions relevant to PSP. I decided to use a balanced random forest machine learning algorithm (RF) with a threshold-moving approach for optimal object classification and interpretability. Brain regions were grouped together based on tau distribution and morphology, resulting in 4 separate tau classifiers for cortical regions, striatum, putamen and subthalamic nucleus, and dentate nucleus. I optimised the algorithm to quantify tau pathological hallmarks of PSP which include 'coiled bodies' (CB), 'neurofibrillary tangles' (NFT), 'tufted astrocytes' (TA) and ‘tau fragments’ (TF). A successful pipeline should have comparable performance in the held-out test set to the expert pathologists.

This work was performed in collaboration with Annelies Quaegebeur (Senior Clinical Research Associate and Consultant Neuropathologist at the University of Cambridge and Cambridge University Hospital NHS Foundation Trust) who assisted with tau annotation and gave advice on pathology; Sanne Kaalund (post-doctoral researcher with James Rowe and the Cambridge brain bank) who extracted brain slices, arranged slide staining and scanning; and Eric Hidari (post-doctoral researcher in the department of chemistry) who assisted with brain slide segmentation.

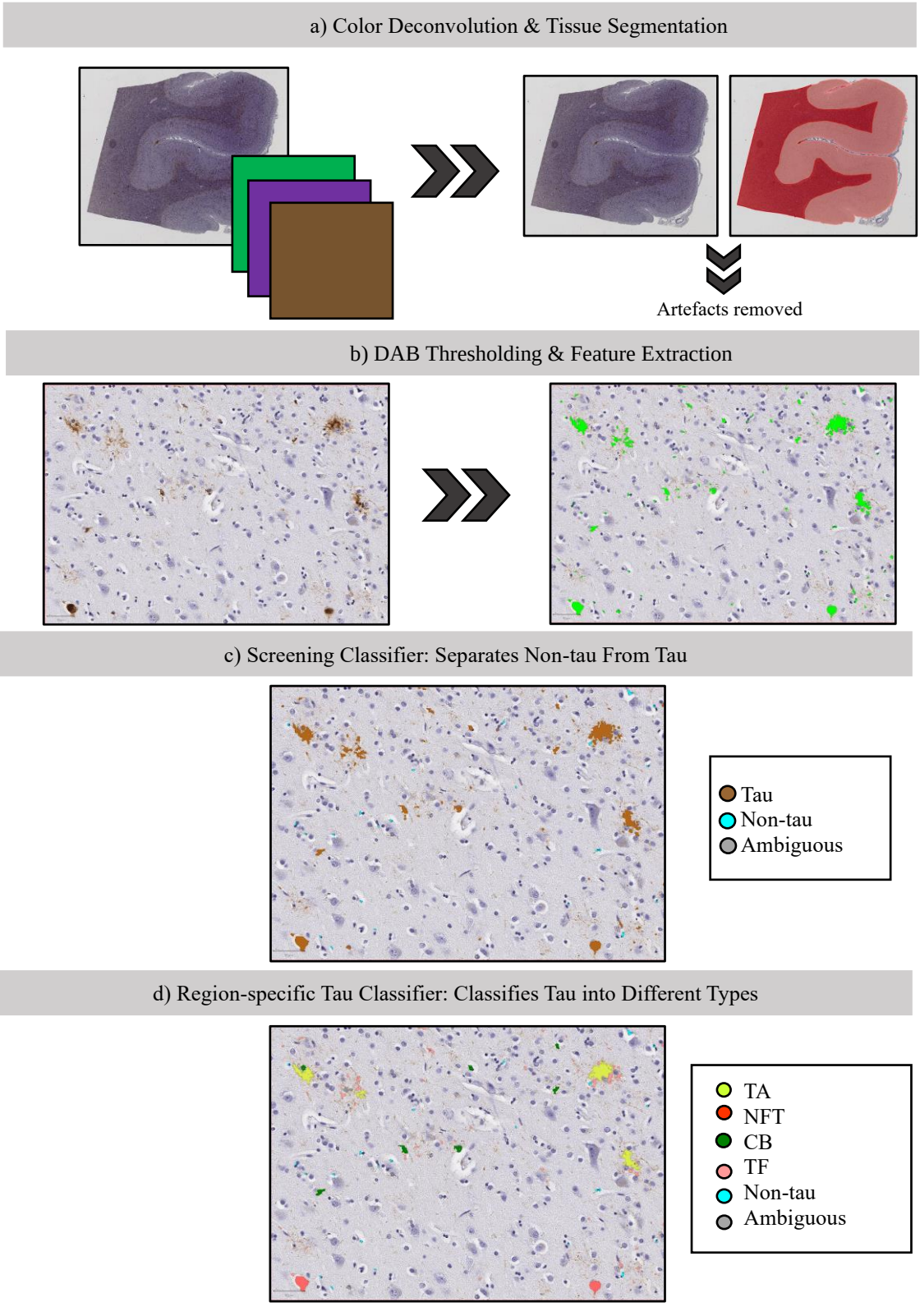



Figure 4.1: Tau pipeline overview. a) In Qupath, a whole slide image undergoes colour deconvolution and segmentation of region of interest. Artefacts are manually removed. b) DAB thresholding is performed to detect tau objects (in green) and features are extracted. c) The screening classifier separates non-tau artefacts from tau objects and d) tau objects are classified into different tau types. Final slide checking to ensure accurate results before subsequent analysis.

## 4.2. Materials and methods

Table 4.1: Demographic, clinical and region-specific severity rating, overall PSP stage and clinical diagnosis of donor participants in the study. Brain slides from each participant used for training and as held-out test set are indicated. Not applicable (N/A) where data is not available in the database.

| Subject | GP | STN | STR | PF | DN | OC | Stage | Clinical diagnosis | Gender | Age at death (years) | Disease duration (years) | Training | Held-out |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 2 | 0 | 1 | 0 | 2 | prob. PSP-RS | Female | 76.4 | 8.75 | OC | - |
| 2 | 2 | 2 | 2 | 1 | 1 | 0 | 3 | prob. PSP-RS | Male | 74.5 | 6 | OC | - |
| 3 | 3 | 3 | 3 | 3 | 2 | 0 | 4 | poss. PSP-CBS | Male | 78.8 | 5.75 | - | DN |
| 4 | 2 | 3 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 80.8 | 11.92 | OC | - |
| 5 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Male | 71 | 5.42 | - | BG |
| 6 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | s.o. PSP-CBS | Male | 73.9 | 4 | - | BG |
| 7 | NaN | NaN | NaN | NaN | NaN | NaN | 5 | poss. PSP-CBS | Male | 78 | NaN | - | DN |
| 8 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Male | 76.2 | 3.87 | PF, 1°M, BG | PM |
| 9 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | poss. PSP-SL | Female | 78.4 | 8.83 | BG | - |
| 10 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Female | 69.9 | 5.33 | BG, DN | - |
| 11 | 3 | 3 | 3 | 1 | 3 | 1 | 5 | poss. PSP-CBS | Female | 78.9 | 5 | BG | - |
| 12 | 2 | 3 | 2 | 3 | 2 | 1 | 5 | prob. PSP-RS | Male | 71.5 | 5.17 | DN | - |
| 13 | 2 | 2 | 2 | 2 | 2 | 1 | 5 | prob. PSP-RS | Female | 74.7 | 6.58 | DN | - |
| 14 | 3 | 3 | 2 | 2 | 3 | 2 | 6 | s.o. PSP-CBS | Female | 80.5 | 8.42 | PM | PF |

Severity rating for each brain region includes 0 = absence, 1 = mild, 2 = moderate, 3 = severe. For clinical diagnosis of PSP participants, *prob.* probable, *poss.* possible, s.o. suggestive of, *RS* Richardson syndrome, *CBS* predominant corticobasal syndrome, *SL* predominant speech and language disorder. *PF* pre-frontal, *1°M* Primary motor, *PM* Pre-motor, *OC* Occipital, *BG* Basal ganglia, *DN* Dentate nucleus.

### 4.2.1. Donors and brain regions

A total of 19 formalin-fixed paraffin embedded slides were obtained from 14 brains (1-5 slides per brain, median = 1.5, IQR = 1) donated by patients with a clinical and pathological diagnosis of Progressive Supranuclear Palsy (PSP) (Table 4.1). A total of 13 slides were used for model development, 6 cortical slides (1 pre-motor, 1 pre-frontal, 1 primary motor and 3 occipital slides), 3 dentate nucleus slides and 4 basal ganglia slides. Occipital slides were

used to only train the screening classifier due to heavy artefacts. The held-out test set consisted of 6 extra slides (2 cortical, 2 basal ganglia and 2 dentate nucleus slides). Training and held-out test slides were annotated by a trained expert (TP), and a neuropathologist (AQ) independently annotated the held-out test slides to calculate the inter-rater reliability.

### 4.2.2. DAB thresholding and feature extraction

Table 4.2: Haralick and morphological features extracted from detected objects and used for tau classification.

| | **Features** |
|---|---|
| Haralick features | Angular second moment (F0) |
| | Contrast (F1) |
| | Correlation (F2) |
| | Sum of squares (F3) |
| | Inverse difference moment (F4) |
| | Sum average (F5) |
| | Sum variance (F6) |
| | Sum entropy (F7) |
| | Entropy (F8) |
| | Difference variance (F9) |
| | Difference entropy (F10) |
| | Information measure of correlation 1 (F11) |
| | Information measure of correlation 2 (F12) |
| Morphology | Area |
| | Circularity |
| | Length |
| | Maximum diameter |
| | Minimum diameter |
| | Solidity |

A thresholder tool in QuPath software (Bankhead et al., 2017) was applied to the DAB channel to detect tau objects (*resolution = high, pre-filter = Gaussian, smoothing sigma = 0, threshold = 0.25, minimum object size = 5mm$^2$).* Areas with DAB intensity above the threshold were labelled as tau objects. Optimal parameters of the thresholder were obtained from visual inspection to maximise the detection of tau and minimise the detection of noise and artefacts.

To reduce the creation of artefacts resulting from bleeding of digital stains between the haematoxylin and DAB channels, we applied an initial screening classifier. This is a random forest classifier trained on all extracted features to separate non-tau from tau objects. Non-tau objects include artefacts from slide preparation, and brown biological elements such as iron granules and lipofuscin.

In total, 54 features were extracted from each tau object. These comprised 35 intensity features, where 5 features (*minimum, maximum, mean, median and standard deviation*) were calculated from 7 channels (*red, green, blue, DAB, haematoxylin, brightness, and saturation*), 6 morphological features and 13 Haralick features from the DAB channel were also computed for textural information (see Table 4.2).

### 4.2.3. Training set

An important step of training a machine learning algorithm is to create a training set. This training set contains manually labelled data from which the machine learning algorithm 'learns'. To create an equal sampling area for each training slide, a grid view was used (*grid size = 250 x 250 $\mu m$*). Each tau object labelled by DAB thresholding was manually labelled as belonging to one of the five classes ('coiled body' (CB), 'neurofibrillary tangle' (NFT), 'tufted astrocyte' (TA), 'tau fragments' (TF), and 'non-tau').

A screening classifier was trained on 9,827 tau and 12,006 non-tau objects annotated from cortical and basal ganglia slides (see Figure 4.2). The screening classifier employed a random forest algorithm using the features in Table 4.2.

For the cortical tau classifier, training objects were sampled from boxes defined over areas of high tau burden, yielding 3,954 objects (661 CB, 126 NFT, 254 TA, 2913 TF). For basal ganglia and the dentate nucleus, 4-grid boxes with 1-grid spacing between the boxes were drawn to cover the entire area for sampling. The tau classifier for the putamen was trained on 3,699 tau objects (335 CB, 48 NFT, 200 TA, 3116 TF), and the tau classifier for the subthalamic nucleus and globus pallidus was trained on 13,686 tau objects (601 CB, 97 NFT, 12988 TF). The tau classifier for the dentate nucleus was trained on 2,186 tau objects (147 CB, 234 NFT, 1805 TF). The tau classifiers for the subthalamic nucleus and globus pallidus, and dentate nucleus were not trained to detect TA as they are very rare in these regions, unlike in the putamen and the cortex. Sample size checks were carried out to ensure sufficient data for training the algorithm.

### 4.2.4. Held-out test set

Two slides from each of the cortical, basal ganglia and dentate nucleus slides were randomly selected as held-out test slides and annotated by a trained expert (TP) and a neuropathologist (AQ). Cohen's kappa was used to assess the inter-rater reliability alongside classification performance against the trained expert. In total, 5754 objects were annotated for cortical slides (296 CB, 78 NFT, 237 TA, 1761 TF, 3382 non-tau). For the basal ganglia, 6528 objects were annotated (153 CB, 21 NFT, 2795 TF, 44 TA, 3515 non-tau), with 2207 objects in the globus pallidus, 2199 objects in putamen, 2122 objects in the subthalamic nucleus. For dentate nucleus, 2280 objects were annotated (18 CB, 26 NFT, 844 TF, 1392 non-tau).

## 4.2.5. Model development

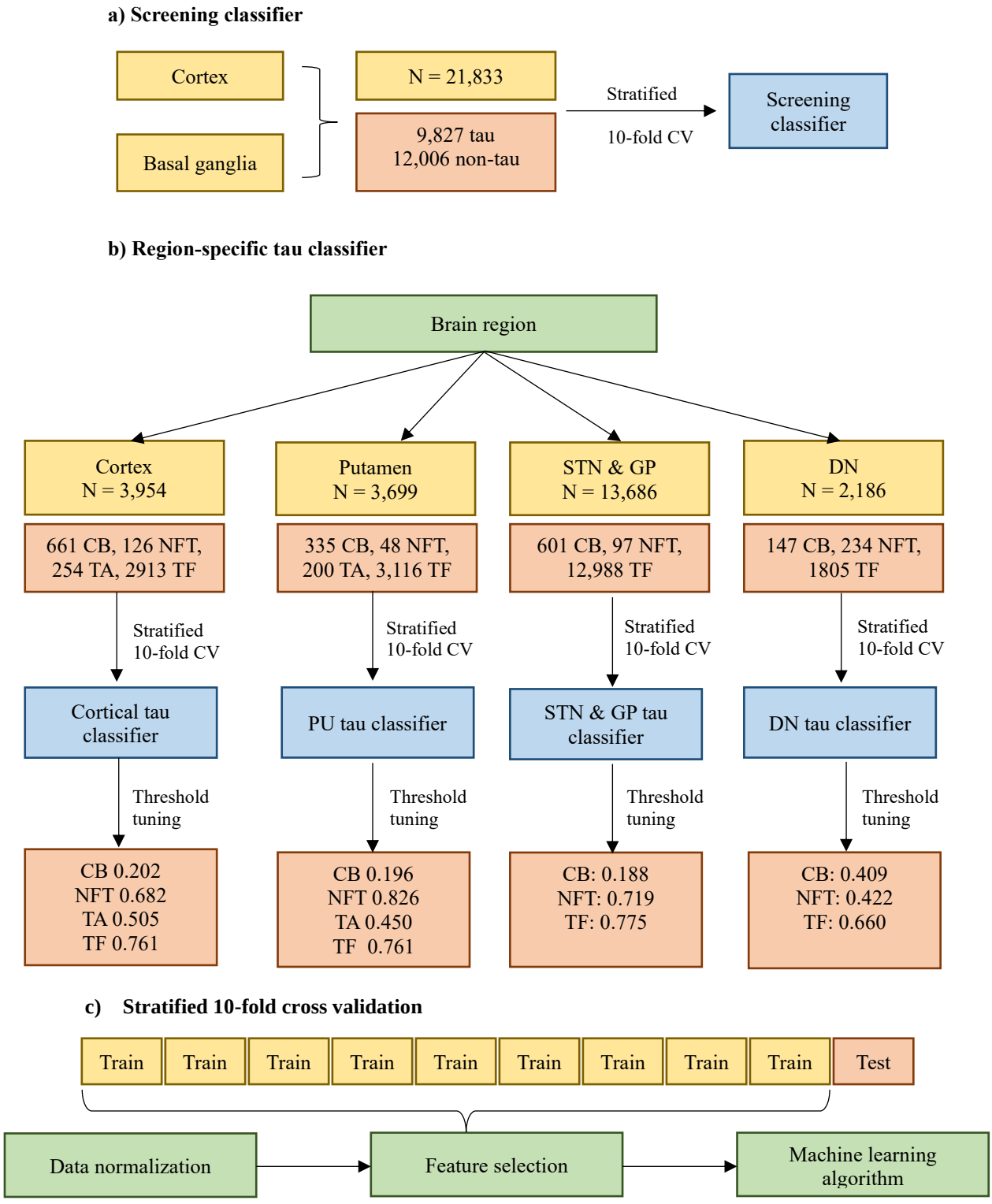



Figure 4.2: Schematic diagram showing annotated data and hyper-parameter tuning step for the a) screening classifier and b) region-specific tau classifiers. c) For each loop through the stratified 10-fold cross validation, the following steps were carried out: data normalization, hyper-parameter tuning of feature recursive elimination with a random forest and the balanced random forest algorithm for tau classification. PU Putamen, STN & GP Subthalamic nucleus and globus pallidus, DN Dentate nucleus.

Annotated objects from all slides of the same brain grouping were pooled together, yielding a dataset. In this study, there were 4 regional groupings: cortex, putamen, globus pallidus and subthalamic nucleus, and dentate nucleus. The dataset was then standardised (*mean = 0, SD =1*) and 10-fold stratified cross validation was used to train the region-specific cell classifiers. Balanced random forest algorithm was implemented for the tau classification pipeline. During the training phase (Figure 4.2), hyper-parameters of the feature selection step using recursive feature elimination with random forest and the balanced random forest for tau classification were tuned using a random search. The balanced random forest was optimised based on the mean area under the 4 precision-recall curves (PR-AUC) using a one-vs-rest approach (TA vs rest, CB vs rest, NFT vs rest, TF vs rest).

Using the optimised hyper-parameters, class-specific thresholds were tuned to tackle class imbalance. Brain regions with similar tau morphology and distribution were grouped together where class-specific thresholds were tuned separately for each brain grouping.

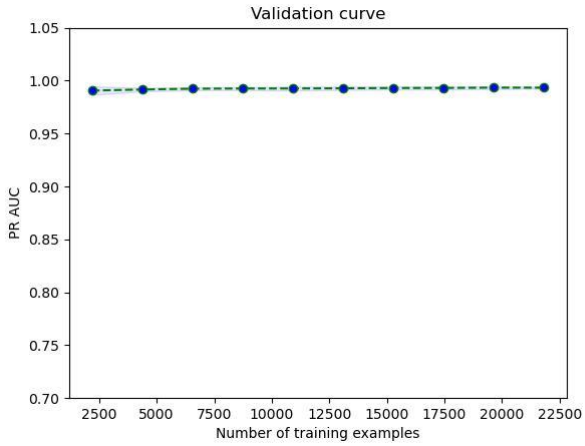
After classification, the precision, recall, macro F1-score and confusion matrix of the model were assessed. The model was then applied to the held-out test set to evaluate its performance and generalisability.

## 4.3. Results

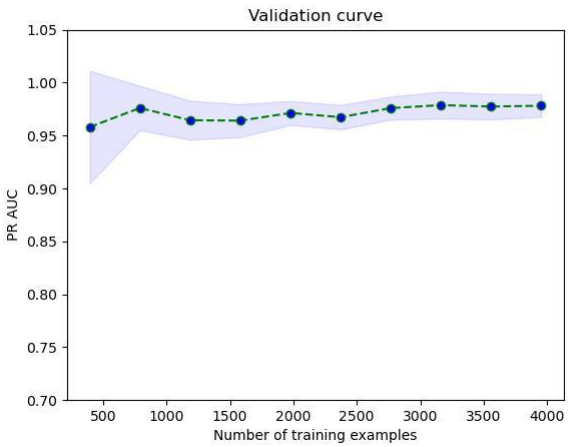
### 4.3.1. Sample size check

First, we ensured there was sufficient data to train the screening and region-specific tau classifiers for optimal results. We plotted the mean PR-AUC from the validation sets in the 10-fold cross validation to check the stability of the score when sampling from 10-100% of the dataset (Figure 4.3). The screening classifier had high PR-AUC of 0.99 (0.9906 - 0.9934) consistently from 10-100% of the dataset. All the tau classifiers showed a similar trend where mean PR-AUC was the lowest when 10% of the data was used and slowly increased as more data was used. The 95% confidence interval of the estimates also became smaller as more data was used to train the classifier. The estimates stabilised after 60-70% of the data was used where mean PR-AUC remained relatively unchanged and 95% confidence intervals stayed small.

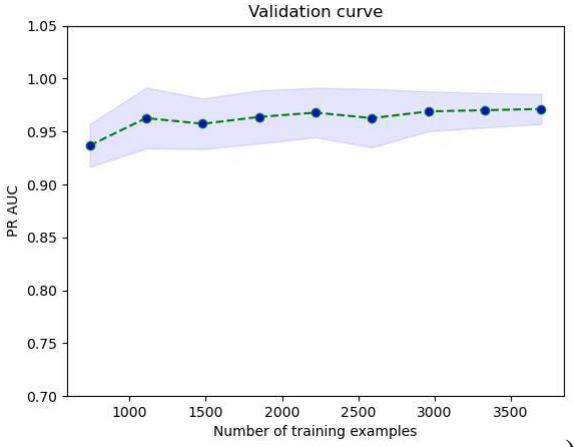
a) Screening tau classifier



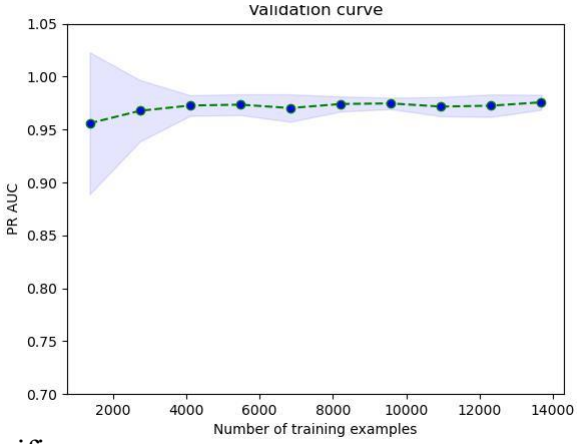

b) Cortical tau classifier




c) Putamen tau classifier




d) STN&GP tau classifier




e) DN tau classifier

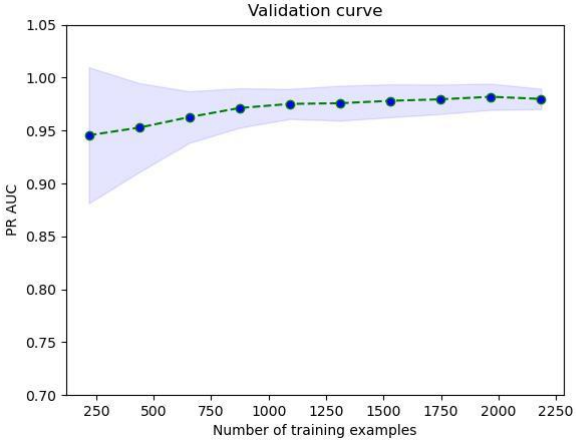



Figure 4.3: Plots showing mean PR-AUC score for each classifier from 10-fold stratified cross validation when 10-100% of the dataset was used. This shows the stability of PR-AUC to inform whether increasing dataset would be useful in training the classifiers. Mean (dot) and 95% confidence intervals (shaded area) are shown.

### 4.3.2. Hyper-parameter tuning

Table 4.3: Hyperparameter values from stratified 10-fold cross validation for the screening classifier and tau classifier for each regional grouping.

| Parameter | Screening | Cortical | Putamen | STN & GP | Dentate nucleus |
|---|---|---|---|---|---|
| N_features_to_select | 46 | 40 | 34 | 34 | 34 |
| Sampling strategy | 'auto' | 'not majority' | 'not majority' | 'not majority' | 'not majority' |
| n_estimator | 600 | 800 | 500 | 500 | 100 |
| min_sample_split | 2 | 2 | 2 | 2 | 2 |
| min_sample_leaf | 2 | 1 | 2 | 2 | 1 |
| max_features | 1 | 0.2 | 0.6 | 0.6 | 0.2 |
| max_depth | None | 10 | 15 | 15 | None |
| max_sample | None | 0.75 | 0.75 | 0.75 | None |

*STN & GP* Subthalamic nucleus and globus pallidus.

All classifiers were optimised for PR-AUC where tau classifiers for different regions yielded different set of optimal hyperparameter values, except for the classifiers for basal ganglia nuclei (Table 4.3). All classifiers achieved PR-AUC scores of over 0.97 (Table 4.4) where the screening classifier achieved the highest PR-AUC of 0.99, and the tau classifier for the subthalamic nucleus and globus pallidus achieved the lowest PR-AUC of 0.97. Tau classifiers for the cortex, putamen, and dentate nucleus achieved similar PR-AUC scores of 0.98. Tau classifiers for non-cortical regions selected 34 from 54 features, while 40 features were selected for the cortical tau classifier and 46 features were selected for the screening classifier from the hyper-parameter tuning step. For feature importance (Figure 4.4), the top 10 most important features for the screening classifier were mainly staining intensities in hematoxylin, red and DAB channels. Tau classifiers for different brain regions showed the same trend where morphological features such as area and diameter of tau objects were the most important, followed by staining intensities and textural features.

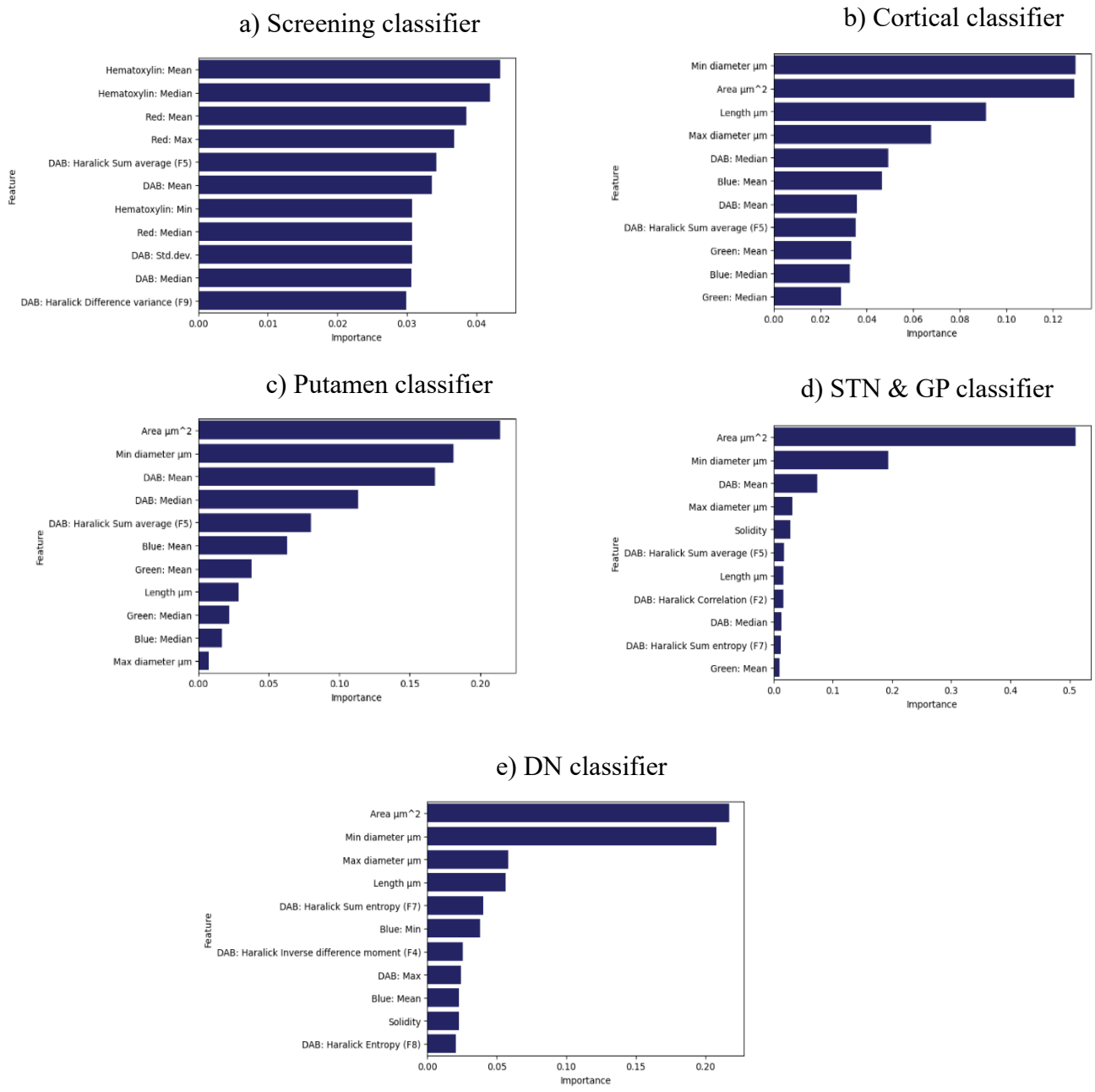



Figure 4.4: Top ten most important features of each classifier from hyper-parameter tuning. a) screening classifier, b) cortical tau classifier, c) tau classifier for putamen, d) tau classifier for subthalamic nucleus and globus pallidus (STN & GP), and e) tau classifier for dentate nucleus (DN).

Table 4.4: Classification performance of tau object identification with no threshold-moving applied, for the screening classifier, and tau classifiers. Classifiers were tuned for area under the precision-recall curve (PR-AUC), where precision, recall and F1-score were calculated. Mean values from cross-validation and standard deviation in brackets are reported.

| Classifier | Precision | Recall | F1-score | PR-AUC |
|---|---|---|---|---|
| Screening | 0.96 (+/-0.044) | 0.96 (+/-0.053) | 0.96 (+/-0.060) | 0.99 (+/-0.011) |
| Cortex | 0.92 (+/-0.037) | 0.92 (+/-0.017) | 0.91 (+/-0.027) | 0.98 (+/-0.010) |
| Putamen | 0.90 (+/-0.046) | 0.86 (+/-0.068) | 0.86 (+/-0.061) | 0.98 (+/-0.015) |
| STN & GP | 0.93 (+/-0.038) | 0.86 (+/-0.056) | 0.87 (+/-0.061) | 0.97 (+/-0.020) |
| DN | 0.96 (+/-0.020) | 0.93 (+/-0.041) | 0.94 (+/-0.029) | 0.98 (+/-0.016) |

*STN & GP* Subthalamic nucleus and globus pallidus, *DN* Dentate nucleus.

Table 4.5: Classification performance of tau object identification with the threshold-moving method applied, for the screening classifier and tau classifiers. Precision, recall and F1-score are calculated where mean values from cross-validation and standard deviation in brackets are reported.

| Classifier | Precision | Recall | F1-score |
|---|---|---|---|
| Screening | 0.96 (+/-0.046) | 0.96 (+/-0.057) | 0.95 (+/-0.064) |
| Cortex | 0.95 (+/-0.025) | 0.95 (+/-0.027) | 0.95 (+/-0.021) |
| Putamen | 0.94 (+/-0.024) | 0.92 (+/-0.049) | 0.93(+/-0.037) |
| STN & GP | 0.95 (+/-0.026) | 0.92 (+/-0.042) | 0.93 (+/-0.027) |
| DN | 0.96 (+/-0.028) | 0.95 (+/-0.026) | 0.95 (+/-0.023) |

*STN & GP* Subthalamic nucleus and globus pallidus, *DN* Dentate nucleus.

Table 4.6: Class-specific thresholds of tau classifiers for the cortex, putamen, subthalamic nucleus and globus pallidus (STN & GP) and dentate nucleus (DN). Thresholds were optimised for F1-score using a one-vs-rest approach. Not applicable (N/A) is reported where TA is not quantifiable.

| Classifier | CB | F1-score | NFT | F1-score | TA | F1-score | TF | F1-score |
|---|---|---|---|---|---|---|---|---|
| Cortex | 0.20 | 0.95 | 0.68 | 0.91 | 0.51 | 0.96 | 0.76 | 0.99 |
| Putamen | 0.20 | 0.88 | 0.83 | 0.98 | 0.45 | 0.94 | 0.76 | 0.99 |
| STN & GP | 0.19 | 0.90 | 0.72 | 0.95 | N/A | N/A | 0.78 | 1.00 |
| DN | 0.41 | 0.91 | 0.42 | 0.98 | N/A | N/A | 0.66 | 0.99 |

### 4.3.3. Threshold-moving optimisation

After optimising the hyper-parameters for each classifier, the next step was to tune class-specific thresholds for assigning labels to individual tau objects. Using a one-vs-rest approach, the class threshold with the highest F1-score was selected (Table 4.6). For the screening classifier, the threshold for tau (threshold = 0.46; F1-score 0.97) was lower than non-tau (threshold = 0.53; F2-score 0.97). The threshold for TF (threshold=0.76) was the highest for the cortical tau classifier, followed by NFT (threshold = 0.68), TA (threshold = 0.51) and CB class (threshold = 0.20). The class thresholds of the tau classifier for the putamen followed a similar pattern (TA threshold = 0.45, CB threshold = 0.20) but differed in that the NFT threshold (0.83) was higher than the TF class (0.76). The tau classifier for the subthalamic nucleus and globus pallidus and dentate nucleus followed the same trend, where the TF class threshold was highest (STN & GP = 0.78, DN = 0.66), followed by the NFT (STN & GP = 0.72, DN = 0.42), and CB classes (STN & GP = 0.19, DN = 0.41).

We further compared the classification performance of the threshold-moving method to the default method of assigning class labels based on F1-scores. The screening classifier with or without the threshold-moving method performed similarly, with F1-scores of 0.95 and 0.96 respectively. Therefore, the screening classifier without threshold-moving was selected as the final screening model. For tau classifiers, the threshold-moving method improved the mean F1-score and were used in the final models (Table 4.4, Table 4.5).

### 4.3.4. Final models and confusion matrices

Confusion matrices for each classifier are shown in Figure 4.5. The screening classifier achieved high accuracy for both tau (97.75%) and non-tau (93.75%) with minimal misclassification. Tau classification for cortical regions achieved the highest accuracy of 99.17% in classifying TF correctly, followed by TA (96.71%), CB (93.53%) and NFT (89.17%). NFT was misclassified as CB most often (6.67%) while CB was most often misclassified as TF (3.08%). Similarly, tau classification for the putamen achieved the highest accuracy in classifying TF (99.03%) followed by TA (95.21%). However, the classifier misclassified CB most often (accuracy 84.16%) as opposed to NFT (accuracy 89.13%). CB was most wrongly classified as TF (10.87%) but not vice versa. NFT was wrongly classified as either TA (6.52%) or CB (4.35%), but never as TF whilst TA was most often classified as CB (3.72%). For tau classifiers in regions with no TA quantified, they performed best in classifying TF correctly (99.33% for dentate nucleus, 99.59% for subthalamic nucleus and globus pallidus). Tau classification was slightly lower in the subthalamic nucleus and globus pallidus compared to the dentate nucleus in classifying NFT (88.76% vs 96.89%) and CB (86.71% vs 90.07%) correctly. For misclassifications, a similar pattern was seen in both regions where CB was mostly misclassified as TF and NFT, while TF were rarely misclassified. The proportion of objects labelled as ‘Ambiguous’ from each of the tau classifiers was no more than 1 % of tau objects.

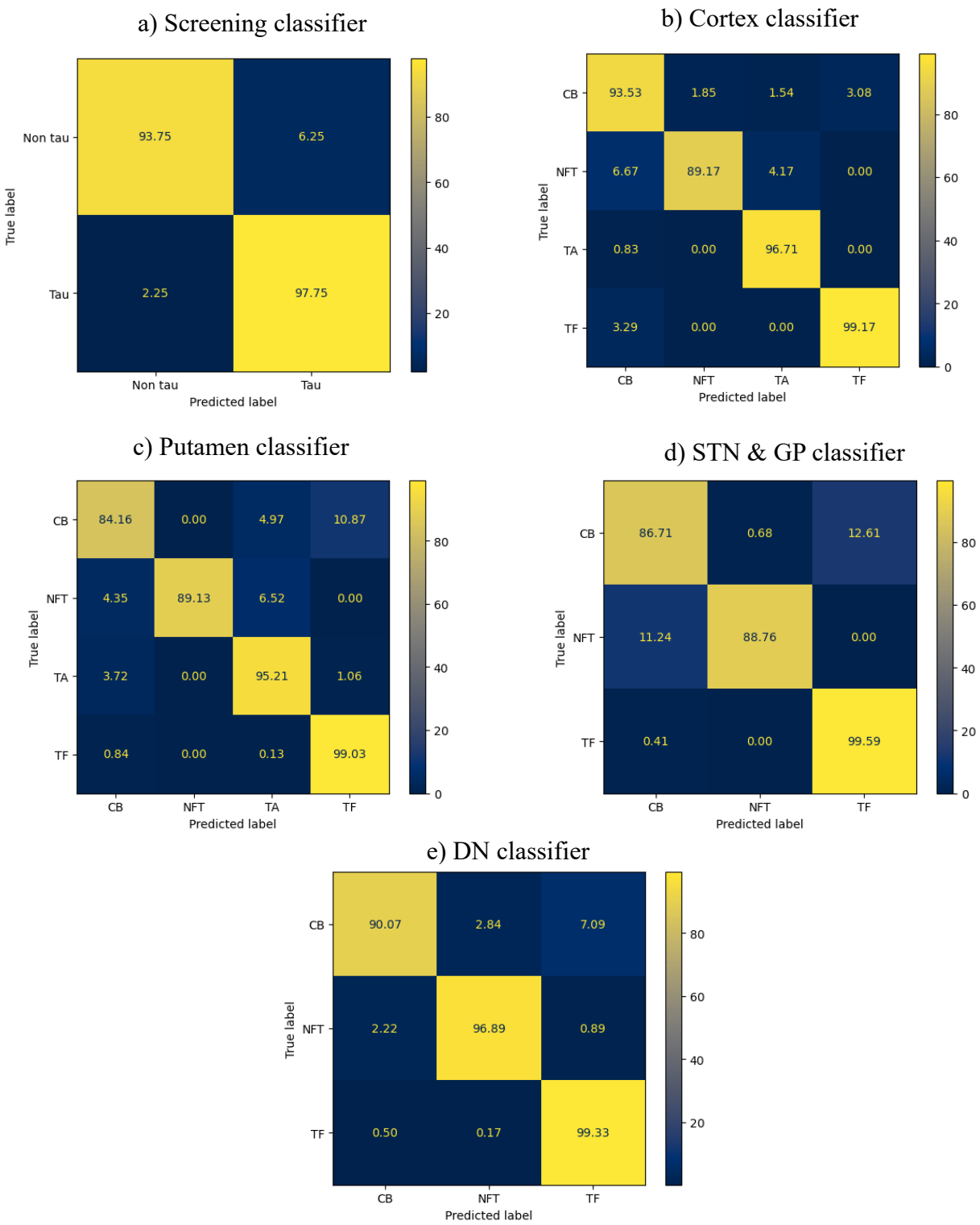



Figure 4.5: Mean confusion matrices from validation set in the 10-fold cross validation for a) screening classifier and tau classifier for b) the cortex, c) putamen, d) subthalamic nucleus and globus pallidus (STN & GP), and e) dentate nucleus (DN).

### 4.3.5. Validation on the held-out test set

Table 4.7: Classification performance on a held-out test set. Precision, recall and F1-score are reported and supplemented with Cohen's kappa to show agreement between the raters (algorithm, rater 1, rater 2).

| Region | Precision | Recall | F1-score | Algorithm & rater 1 | Rater 1 & 2 |
|---|---|---|---|---|---|
| Cortex | 0.98 | 0.96 | 0.97 | 0.94 | 0.96 |
| Putamen | 0.97 | 0.98 | 0.98 | 0.97 | 0.99 |
| Subthalamic nucleus | 0.91 | 0.94 | 0.92 | 0.87 | 1.00 |
| Globus pallidus | 0.95 | 0.96 | 0.96 | 0.93 | 0.99 |
| Dentate nucleus | 0.91 | 0.95 | 0.93 | 0.97 | 1.00 |

From Table 4.7, using rater 1 as the ground truth (the trained expert), the F1-score of the classification performance in each of the brain region ranged from 0.92 to 0.98. The classifier performed best in the putamen, followed by the cortex, globus pallidus, dentate nucleus and subthalamic nucleus. Furthermore, Cohen's kappa indicated that the agreement between rater 1, the algorithm, and rater 2 across brain regions was high, at least 0.87. The agreement between two human raters was higher than the algorithm and rater 1 across all regions, where the smallest difference was by 0.02 in the cortex and putamen, followed by 0.03 difference in dentate nucleus, 0.06 in globus pallidus and 0.13 difference in the subthalamic nucleus, indicating low levels of classification uncertainty in each region. Looking at the confusion matrices (Figure 4.6), the algorithm achieved above 90% accuracy in classifying tau types across all brain regions but struggled more with classifying CB accurately in the cortex (89.44%) and dentate nucleus (83.33%) as they could be mistaken for TF.

Figure 4.7 displays examples of correct classification of tau type-specific aggregates across all brain regions. CB has a coiled-like structure and can appear larger in subcortical structures compared to the cortex. Similarly, NFT is a highly pigmented oval structure and can appear larger in subcortical structures, particularly the subthalamic nucleus, than in the cortex. The dentate nucleus has numerous pre-tangles which are generally more diffuse and granular than NFT and are detected as NFT in the pipeline. Correctly classified TA have star-like tufts of densely packed fibres and appear larger than CB and NFT in general. TF consists of threads

and background tau burden that can often be difficult to associate with a cell. Figure 4.8 illustrates the main misclassifications from the held-out test set where mistakes can occur when tau aggregates have deviant shape from the stereotypical range, meaning their features closely resemble other tau types.

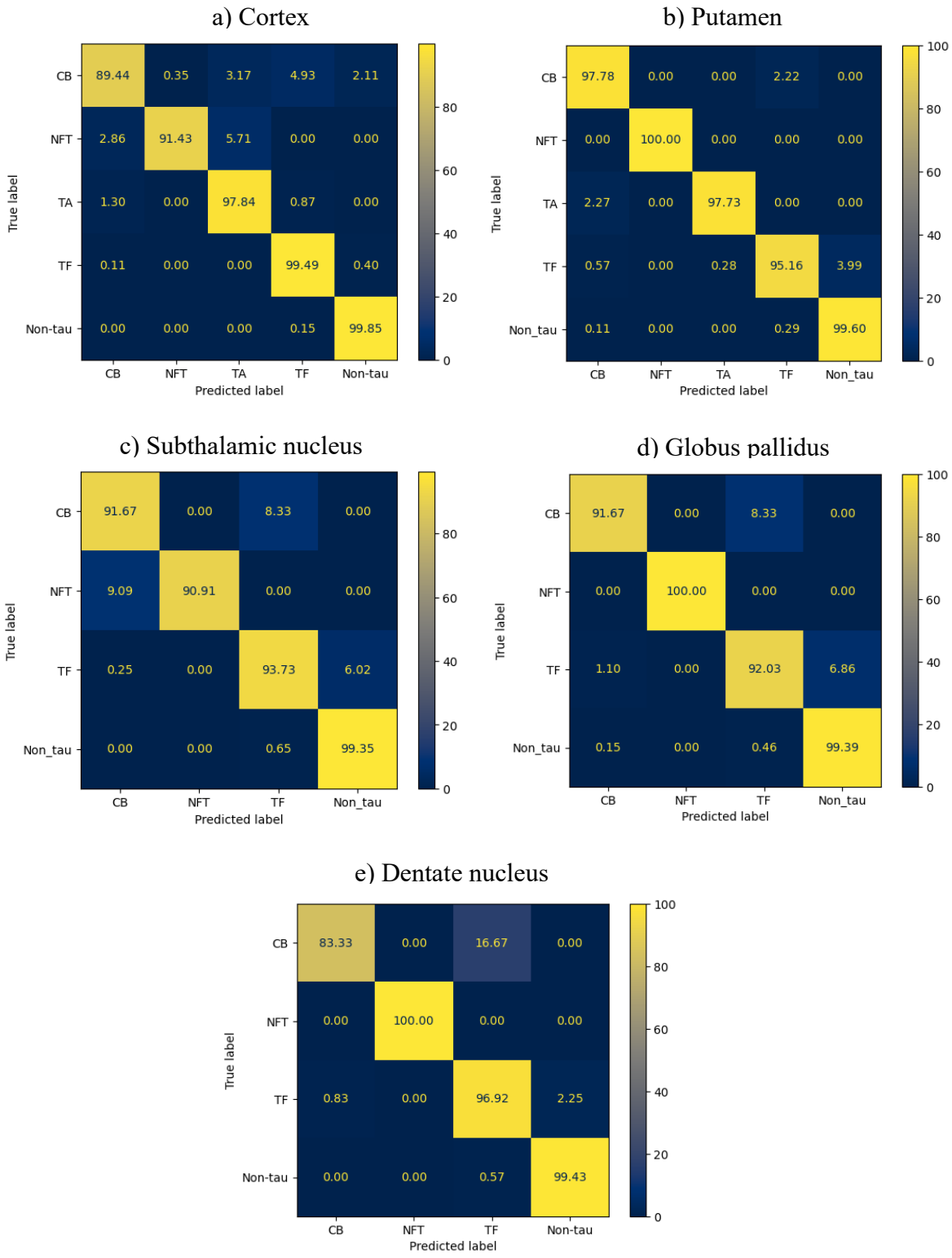



Figure 4.6: Confusion matrices of held-out test set by region a) cortex, b) putamen, c) subthalamic nucleus, d) globus pallidus, e) dentate nucleus.

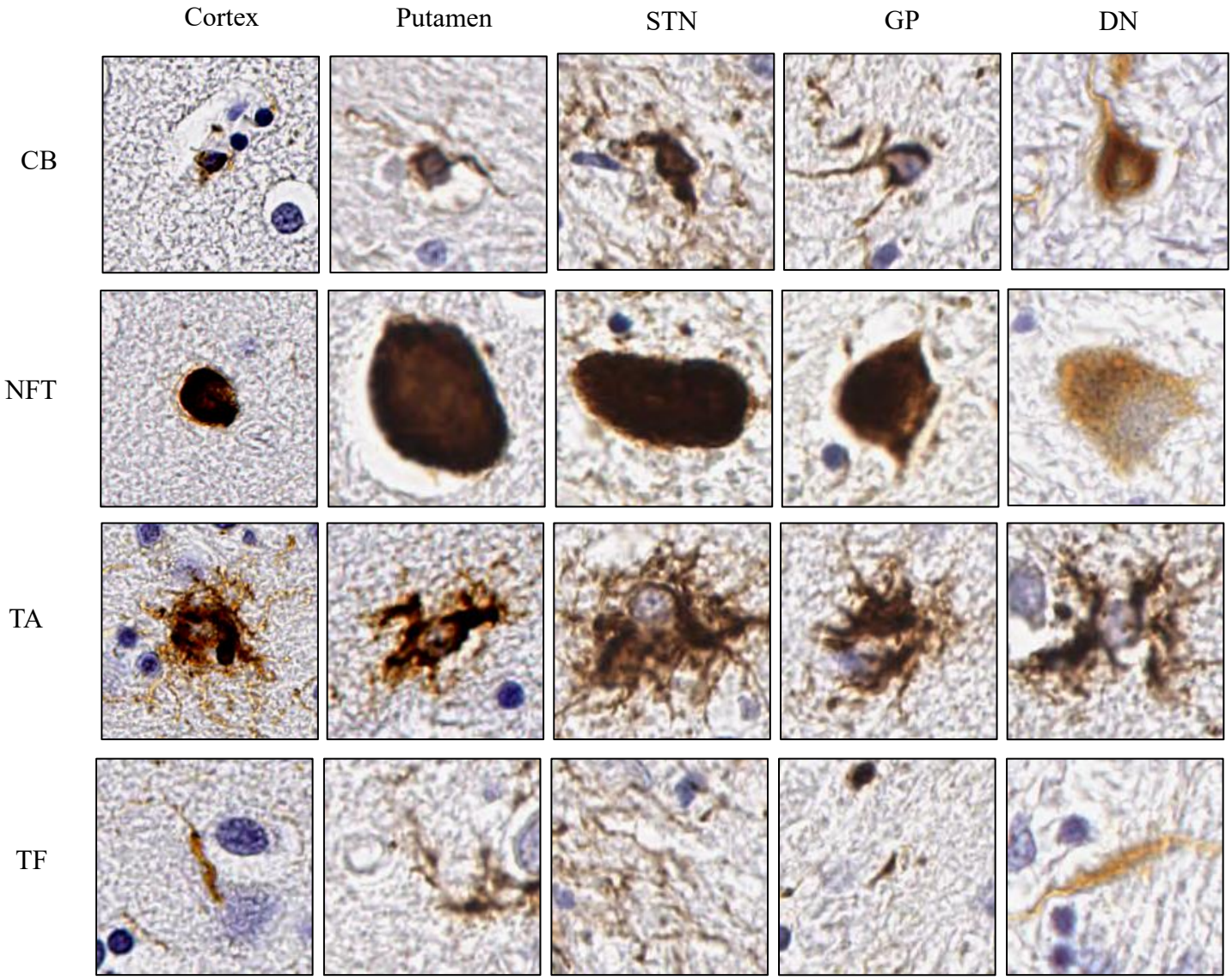



Figure 4.7: Examples of correct classification from the held-out test set for each tau aggregate type from the cortex, putamen, subthalamic nucleus (STN), globus pallidus (GP) and dentate nucleus (DN). All images were cropped 150 x 150 μm window size. TA examples are only drawn from the cortex and putamen. *CB* coiled body; *NFT* neurofibrillary tangle; *TA* tufted astrocyte; *TF* tau fragments.

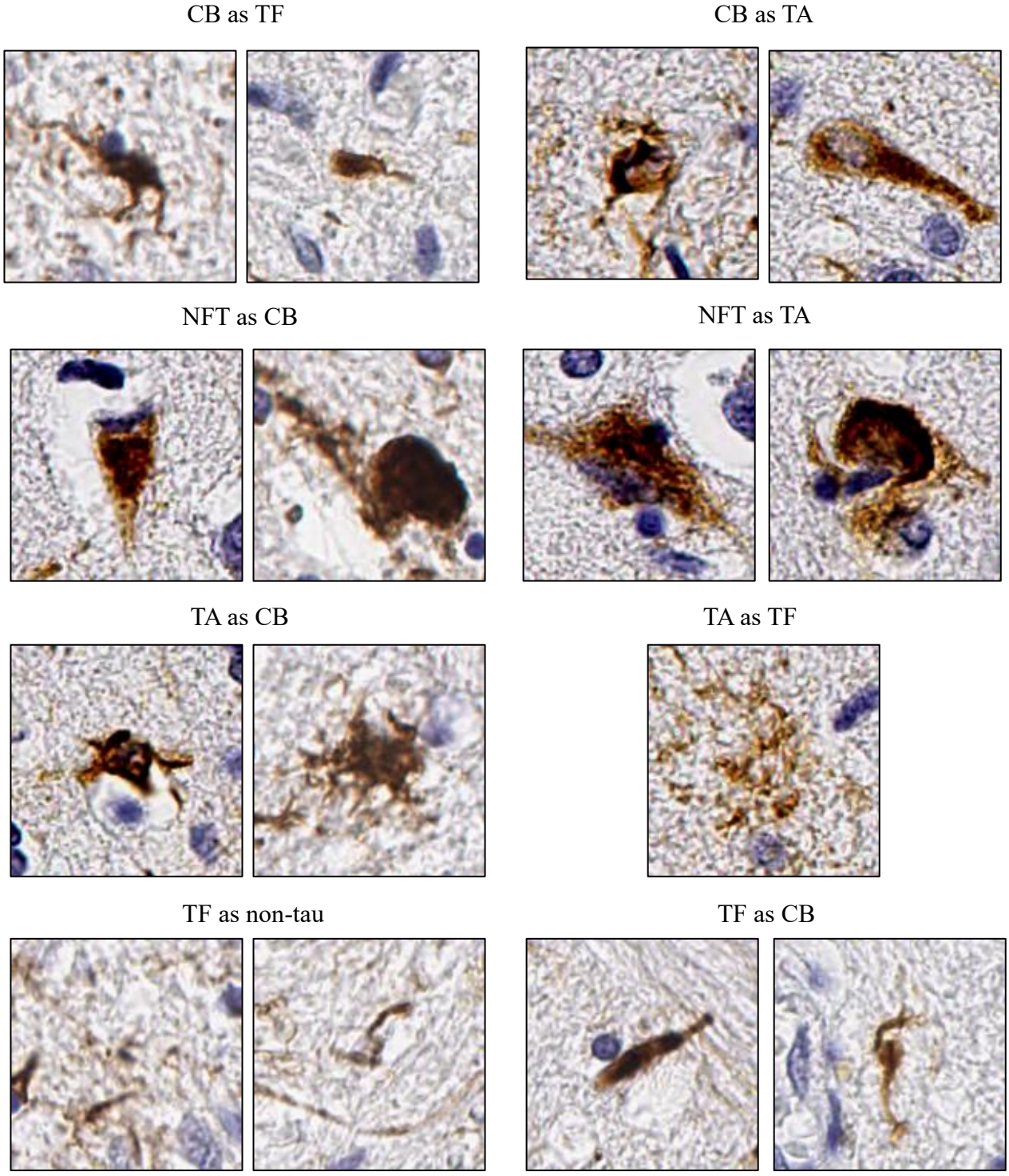



Figure 4.8: Example of possible misclassification observed between tau aggregate types from the held-out test set. Truth label as predicted label is presented. All images were cropped 150 x 150 μm window size. CB coiled body; NFT neurofibrillary tangle; TA tufted astrocyte; TF tau fragments.

## 4.4. Discussion

### 4.4.1. Overview

I have developed a robust and reliable digital pipeline for the quantification of *post mortem* tau pathology in PSP, achieving an accuracy comparable to expert assessment. An optimised balanced random forest machine learning algorithm with a threshold-moving method achieved satisfactory results. The pipeline achieved high classification performance in the held-out test set, F1-score ranging from 0.92 to 0.98 across brain regions in the study. When inspecting the correct and incorrect classifications, it can be observed that tau aggregates of the same type are heterogenous in their morphology. Those that deviate from the stereotypical norm of its own class were the most likely to be misclassified as other tau classes they most closely resembled, which is not surprising given the majority of features weighted highly in the classifier were morphological features. This is largely supported by the observation that CB is rarely ever misclassified as NFT or TA but the confusion between the more morphologically similar CB and TF as non-tau were more frequent.

### 4.4.2. Strengths and limitations

The main issue in project 1 was to do with the data not having sufficient information to distinguish between cell types reliably. The was less of an issue in the current project as tau aggregates of different classes are more distinctive than cells of different classes. Therefore, a balanced random forest with threshold-moving method was sufficient to achieve a high level of classification.

There remain limitations to our study (Coughlin et al., 2022; Koga et al., 2021; Signaevsky et al., 2019). Despite the high accuracy and robustness of the pipeline, it is designed to only classify tau pathologies that are specific to PSP. If the *post mortem* slide has coexisting tauopathies such as Ageing-related Tau Astrogliopathy, Primary Age-related tauopathy or Alzheimer-type neurofibrillary tangle and thread pathology, the pipeline may not yield accurate results because it has not ‘seen’ them before. These coexisting pathologies are not uncommon in PSP but are generally mild in severity so in most cases their impact is minimal (Jecmenica Lukic et al., 2020; Martinez-Lage & Munoz, 1997; Togo et al., 2002).

Moreover, like other pipelines analysing 2-dimensional instances of 3-dimensional tau aggregates, misclassification can occur. For example, it is difficult to identify tau types when two tau objects are superimposed, as this affects their morphology. Moreover, sectioning the slide can influence the size and shape of tau as cutting through the edge or mid-body of a tau object will result in different morphology. In addition, tau morphologies in the brain are not always well defined with clear cut features. Tau objects can be at an early stage of the development where the typical features of a specific tau type are not yet fully visible. This makes them ambiguous, therefore it is very challenging to reliably identify them correctly. These reasons contribute to the subjective nature of classifying tau pathologies. Lastly, the TF class is made up of parts of axonal tau threads, tufted astrocyte processes and other tau fragments. This presents a challenge to a truly accurate quantification, since a large proportion of these fragments will be associated with larger tau inclusions.

The main strengths of our pipeline are its versatility permitting accurate assessment in multiple brain areas, and scalability allowing assessment across a large number of brain regions and a large number of subjects. The accuracy of the method was similar for neuronal and glial tau pathology densities. We have gone beyond former proof of concept studies, which generally include a small subset of brain regions (Coughlin et al., 2022; Koga et al., 2021; Signaevsky et al., 2019). We deliberately applied the machine learning algorithm to both cortical and subcortical structures. This is particularly important in PSP which affects both cortical and basal ganglia regions.

With this pipeline, more detailed information can be extracted beyond the current semi-quantitative method of grading tau pathology in post-mortem brains. It could be useful for mechanistic studies in uncovering more subtle pathological phenomena and for post-mortem diagnosis, as the pipeline could aid pathologists by providing fast and rich information.

# Chapter 5: Testing tau quantification pipeline beyond the held-out test set

## 5.1. Introduction

In PSP, the distribution of tau type-specific aggregates have been well characterised by the current PSP staging scheme (Kovacs et al., 2020). The PSP staging scheme has also been validated in an independent cohort where PSP stage was related to clinical severity prior to death (Briggs et al., 2021). Neuropathological examination using a semi-quantitative approach is the current standard method for examining neuropathology in *post mortem* brains, but it has limitations (Scheltens & Rockwood, 2011). Using ordinal measure of stage severity to reflect tau burden may also have suboptimal clinicopathologic predictive power as it may mask the subtle pathological changes, which is particularly relevant for studies of disease mechanisms (Marx et al., 2022; Walker et al., 2017).

No study has yet compared semi-quantitative to quantitative approach in quantifying tau burden in PSP across multiple brain regions. As PSP staging system is the current standard, it can act as the ground truth, where the degree of agreement will indicate how robust the pipeline is against a large number of novel slides which may contain out-of-sample objects. For additional validation, it will be important to compare the clinicopathologic predictive power of PSP stage to digitally quantified tau of clinical severity.

### 5.1.1. The present study

Here, I used the tau quantification pipeline to investigate tau pathology in *post mortem* slides from PSP and control participants. I first applied the pipeline to histology slides from both groups, where total tau was quantified from control participants, and 4 tau objects (‘neurofibrillary tangles’ (NFT), ‘coiled bodies’ (CB), ‘tufted astrocytes’ (TA), and ‘tau fragments’ (TF)) were quantified for PSP participants.

My objectives are to:

1. Assess the correlation between tau type-specific burden to understand the relationship between tau types across three spatial levels (all brain regions, cortical regions, and subcortical regions) in PSP donors.
    - I predict that tau burden will be more severe in subcortical structures than cortical structures, where occipital region should have less tau burden than other cortical regions.
    - I predict that TF, CB and NFT density will generally be higher than TA density, but TA density is likely to be higher than other types in the cortex.
2. Validate tau burden estimates from the pipeline, correlating tau burden against the PSP pathology staging scheme.
    - There will be at least moderate correlation between digitally quantified tau burden and PSP pathology staging.
3. Assess whether neuropathological severity (PSP stage, digitally quantified tau burden) predicts clinical severity (PSP rating scale, PSPRS) score in PSP donors.
    - Digitally quantified tau burden is hypothesised to have higher predictive power of clinical severity than PSP stage.
4. Compare total tau burden estimates in *post mortem* slides from control and PSP donors to assess the pipeline's robustness when there should be no or minimal tau.
    - PSP donors should have higher tau burden than control donors across all brain regions, specifically in subcortical regions.

## 5.2. Materials and method

Table 5.1: Demographic, clinical and region-specific severity rating, overall PSP stage and clinical diagnosis of donor participants in the study. Not applicable (N/A) where data is not available in the database.

| Subject | GP | STN | STR | PF | DN | OC | Stage | Clinical diagnosis | Gender | Age at death (years) | Disease duration (years) | Last PSPRS Total | PSPRS to death (years) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 2 | 2 | 0 | 1 | 0 | 2 | prob. PSP-RS | Female | 76 | 8.75 | 63 | 0.32 |
| 2 | 3 | 3 | 2 | 0 | 1 | 0 | 2 | poss. PSP-PGF | Male | 75 | 4.62 | 26 | 0.52 |
| 3 | 2 | 2 | 2 | 1 | 1 | 0 | 3 | prob. PSP-RS | Female | 55 | 5.5 | 53 | 0.41 |
| 4 | 2 | 2 | 2 | 1 | 1 | 0 | 3 | prob. PSP-RS | Male | 74 | 6 | 45 | 0.67 |
| 5 | N/A | N/A | N/A | N/A | N/A | N/A | 3 | prob. PSP-RS | Male | 72 | N/A | N/A | N/A |
| 6 | 2 | 3 | 2 | 1 | N/A | 0 | 4 | prob. PSP-RS | Female | 65 | 13.92 | 54 | 0.75 |
| 7 | 3 | 3 | 3 | 3 | 2 | 0 | 4 | poss. PSP-CBS | Male | 78 | 5.75 | 43 | 0.69 |
| 8 | 2 | 2 | 2 | 1 | 2 | 0 | 4 | poss. PSP-CBS | Female | 79 | 3.42 | 49 | 0.74 |
| 9 | 3 | 3 | 2 | 2 | 3 | 0 | 4 | prob. PSP-RS | Male | 77 | 6.33 | 55 | 0.74 |
| 10 | 2 | 2 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 78 | 5.33 | 54 | 0.56 |
| 11 | 2 | 3 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 80 | 6.5 | 62 | 0.11 |
| 12 | 2 | 3 | 3 | 1 | 2 | 0 | 4 | prob. PSP-RS | Female | 71 | 4.58 | 45 | 0.13 |
| 13 | 3 | N/A | 2 | 1 | 2 | 0 | 4 | poss. PSP-CBS | Female | 75 | 2.83 | N/A | N/A |
| 14 | 2 | N/A | 2 | 2 | 3 | 0 | 4 | prob. PSP-RS | Male | 64 | 5.08 | 38 | 1.09 |
| 15 | 2 | 3 | 2 | 1 | 2 | 0 | 4 | prob. PSP-RS | Male | 80 | 11.92 | 76 | 1.3 |
| 16 | 2 | 3 | 2 | 3 | 2 | 1 | 5 | prob. PSP-RS | Male | 71 | 5.17 | 38 | 2.21 |
| 17 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | poss. PSP-SL | Female | 78 | 8.83 | 72 | 1.94 |
| 18 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Male | 63 | 8.83 | 62 | 2.44 |
| 19 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Male | 76 | 3.87 | 51 | 0.42 |
| 20 | 2 | 2 | 2 | 2 | 2 | 1 | 5 | prob. PSP-RS | Female | 74 | 6.58 | 58 | 2.66 |
| 21 | 2 | N/A | 2 | 2 | 2 | 1 | 5 | prob. PSP-RS | Male | 88 | 5.08 | 53 | 0.3 |
| 22 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Female | 69 | 5.33 | 51 | 0.36 |
| 23 | 2 | N/A | 2 | 2 | 3 | 1 | 5 | prob. PSP-RS | Female | 71 | 6.17 | 60 | 0.05 |
| 24 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Male | 71 | 5.42 | 43 | 1.89 |
| 25 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | poss. PSP-CBS | Male | 73 | 4 | N/A | N/A |
| 26 | 3 | 3 | 3 | 1 | 3 | 1 | 5 | poss. PSP-CBS | Female | 78 | 5 | 48 | 0.18 |
| 27 | 3 | 3 | 3 | 2 | 3 | 1 | 5 | prob. PSP-RS | Male | 84 | 8.75 | 59 | 2.42 |
| 28 | 3 | 3 | 3 | 3 | 3 | 1 | 5 | prob. PSP-RS | Female | 84 | 4.25 | 67 | 0.78 |
| 29 | 3 | 3 | 3 | 3 | 2 | 1 | 5 | prob. PSP-RS | Female | 78 | 16.75 | 52 | 0.99 |
| 30 | N/A | N/A | N/A | N/A | N/A | N/A | 5 | poss. PSP-CBS | Male | 78 | N/A | N/A | N/A |
| 31 | 3 | 3 | 2 | 2 | 3 | 2 | 6 | poss. PSP-CBS | Female | 80 | 8.42 | 73 | 0.23 |
| 32 | 3 | 3 | 3 | 3 | 3 | 2 | 6 | prob. PSP-F | Male | 75 | 8.42 | 81 | 0.76 |

Severity rating for each brain region includes 0 = absence, 1 = mild, 2 = moderate, 3 = severe. For clinical diagnosis of PSP participants, *prob.* probable, *poss.* possible, s.o.

suggestive of, *RS* Richardson syndrome, *CBS* predominant corticobasal syndrome, *SL* predominant speech and language disorder. *PF* pre-frontal, *1°M* Primary motor, *PM* Pre-motor, *OC* Occipital, *BG* Basal ganglia, *DN* Dentate nucleus.

Table 5.2: Demographic and number of slides available from healthy control donor participants in the study.

| Subject | Gender | Age at death (years) | Braak stage | No. of slides |
|---|---|---|---|---|
| 1 | Male | 87 | I | 3 |
| 2 | Female | 70 | I | 3 |
| 3 | Female | 72 | II | 4 |
| 4 | Male | 74 | II | 3 |
| 5 | Female | 70 | II | 4 |
| 6 | Male | 68 | II | 3 |
| 7 | Female | 84 | III | 1 |
| 8 | Male | 82 | III | 2 |

Table 5.3: Brain regions available from each control donor participant in the study.

| Subject | Pre-frontal | Temporal | Parietal | Occipital | Cingulate | Basal ganglia |
|---|---|---|---|---|---|---|
| 1 | ✓ | ✓ | ✓ | - | - | - |
| 2 | ✓ | ✓ | ✓ | - | - | - |
| 3 | ✓ | ✓ | ✓ | - | ✓ | |
| 4 | ✓ | ✓ | - | ✓ | - | - |
| 5 | ✓ | ✓ | - | ✓ | - | ✓ |
| 6 | ✓ | ✓ | ✓ | - | - | - |
| 7 | - | - | - | ✓ | - | - |
| 8 | - | ✓ | - | ✓ | - | - |

### 5.2.1. Donors and brain regions

A total of 240 formalin-fixed, paraffin embedded slides were obtained from 32 brains (2-10 slides per brain, median = 8.5, IQR = 6) donated by patients with a clinical and pathological diagnosis of Progressive Supranuclear Palsy (PSP) (Table 5.1). The slides included 185 cortical slides (29 pre-frontal, 21 premotor, 20 primary motor, 22 primary somatosensory, 23 temporal, 20 parietal, 28 occipital, 22 cingulate), 25 basal ganglia and 30 cerebellar (dentate nucleus) slides. Of the 240 slides, 13 slides were used for model development and 6 as a held-out test set (see in chapter 2). Following pipeline development, 227 novel slides were used for validation against the PSP staging scheme (Kovacs et al., 2020) and all slides were used for further analyses. A total of 23 slides from 8 control donors (Table 5.2, Table 5.3) with no known history of neurodegenerative diseases were used to compare to PSP slides to further validate the tau quantification pipeline.

### 5.2.2. Tau quantification

The four types of tau quantified in PSP brains were CB, NFT, TA and TF. This enabled the calculation of total tau (all tau types) and tau hallmarks (all tau types excluding TF). Using raw counts of tau quantified, tau density was calculated as the number of tau objects per unit area ($\mu m^2$) of the region quantified. For cortical regions, tau density was quantified in cortical grey matter, while the entire nuclei area was used for basal ganglia and the dentate nucleus.

### 5.2.3. Correlation amongst tau aggregate types and with PSP staging

Polar plots using the plotly package in Python (Plotly Technologies Inc., 2015) were used to show regional tau distribution quantified from the pipeline for both total tau and tau density by tau type. Pearsons's correlation coefficient was used to compare the correlation between tau aggregate types across all brain regions, and separately in cortical and subcortical regions. Subsequently, Spearman's rank correlation coefficient was used to compute the correlation between tau density quantified across regions and PSP stage. Correlations between region-specific tau density and region-specific rating were also computed within regions of the PSP staging scheme.

### 5.2.4. Clinicopathological correlations

For this analysis, we included 28 PSP subjects with available PSPRS scores. Due to the skewness of the tau density distribution, a logarithmic transformation ($log_{10}$) was applied. To investigate the relationship between *post mortem* tau and PSPRS score, Bayesian linear mixed regression models were used, employing the *brms* package in R (version 1.4.1717) (Bürkner, 2017, 2018, 2021). The analysis was first carried out with PSP stage as the predictor, PSPRS total score as the outcome variable, and disease duration and PSPRS to death interval as covariates to establish a baseline relationship between the staging scheme and PSPRS score. The same analysis was repeated with tau density quantified from all regions, and separately from only cortical and subcortical regions as the predictor. A single tau density value from each participant was obtained by first summing total tau count and total brain area from all brain slides considered (all regions, only cortical or subcortical regions), and dividing total tau count by total area.

To test whether tau type-specific burden was more informative of PSPRS score than total tau burden, total tau and tau type-specific models were created for model comparison. To estimate the strength of evidence in favor of the tau type-specific models against the total tau model, we used a standard Bayes Factor (BF) cut-off of 3 to indicate at least moderate evidence (Kruschke, 2011).

In the final model, the strength of regression coefficients were assessed using the Region of Practical Equivalence (ROPE) approach. If 95% of the credible interval (Crl) of the regression coefficient falls completely within the ROPE, then the effect of the parameter would be equivalent to the null value for practical purposes (Kruschke, 2011; Kruschke & Liddell, 2018a).

A Gaussian model family was selected based on the distribution of the data. A weakly informative normal prior *(mean=0, SD =100)* was chosen for the regression coefficients and default priors were used for the intercept *(student-t prior; df=3, mean = 53.5, SD = 12.6)* and the sigma *(student-t prior; df=3, mean =0, scale = 12.6).*

Due to the complexity of our analysis, sensitivity analysis of priors was conducted only to assess the effect of prior choice on neuropathological severity (PSP stage, tau burden) in the

final models. We chose two other weakly informative normal priors, one more informative *(mean = 0, SD = 50)* and the other less informative *(mean=0, SD = 150)* to assess the sensitivity of posterior estimates on the prior choice.

### 5.2.5. Comparison between tau burden in control and PSP participants

Boxplots using seaborn package (Waskom, 2021) in Python were used to compare total tau density quantified from control slides to the corresponding brain region from PSP participants. Visual inspection was carried out in Qupath (Bankhead et al., 2017) to confirm the findings.

## 5.3. Results

### 5.3.1. Tau density across cortical and subcortical regions

Total tau was quantified, and cases were grouped based on their PSP pathology stage (stages 1-6), shown in Figure 5.1. Tau pathology density in subcortical regions was greater than in cortical regions, in keeping with the tau staging system suggesting earlier subcortical involvement. In stage 2, tau accumulation was most prevalent in the subthalamic nucleus and globus pallidus, followed by the dentate nucleus and putamen with minimal tau in cortical regions. In stage 3, there was greater tau pathology in subcortical regions and tau could be seen across multiple cortical regions, especially in the frontal regions, while tau pathology in the occipital lobe was minimal. From stages 4 to 6 tau pathology was greatest in subcortical regions, particularly the subthalamic nucleus and globus pallidus, but the density of tau pathology in the cortical areas increases with each stage, particularly in the frontal lobe.

In the cortex, frontal regions were affected to a greater degree which includes primary motor and pre-motor regions. Temporal and parietal regions showed tau accumulation but to a lesser degree than frontal regions, while the occipital region still showed the least accumulation of tau. Examining tau type-specific density plots (Figure 5.2), the density of tau fragments was higher than other tau types across all PSP stages.

When focusing on individual tau hallmarks (not including TF) across PSP stages, CB density was the most abundant tau type. In subcortical regions, this was followed by NFT density, then TA density, but the pattern is reversed in cortical regions. In stage 2, CB and NFT densities followed the general pattern of total tau accumulation where they were predominantly found in subcortical regions, specifically the globus pallidus and subthalamic nucleus. From stages 3 – 6, CB and NFT appeared in cortical regions and continued to accumulate in subcortical regions. The main cortical regions with high CB and NFT densities included motor and parietal regions, with the least affected area being the occipital region.

TA density was highest in the putamen and could be observed in cortical regions. In contrast to CB and NFT, TA density was the highest in the putamen in stage 2 and was minimal in cortical regions. As the stage progressed, TA density increased predominantly in the putamen and

cortical regions and TA density showed the same pattern as NFT and CB accumulation in cortical regions.

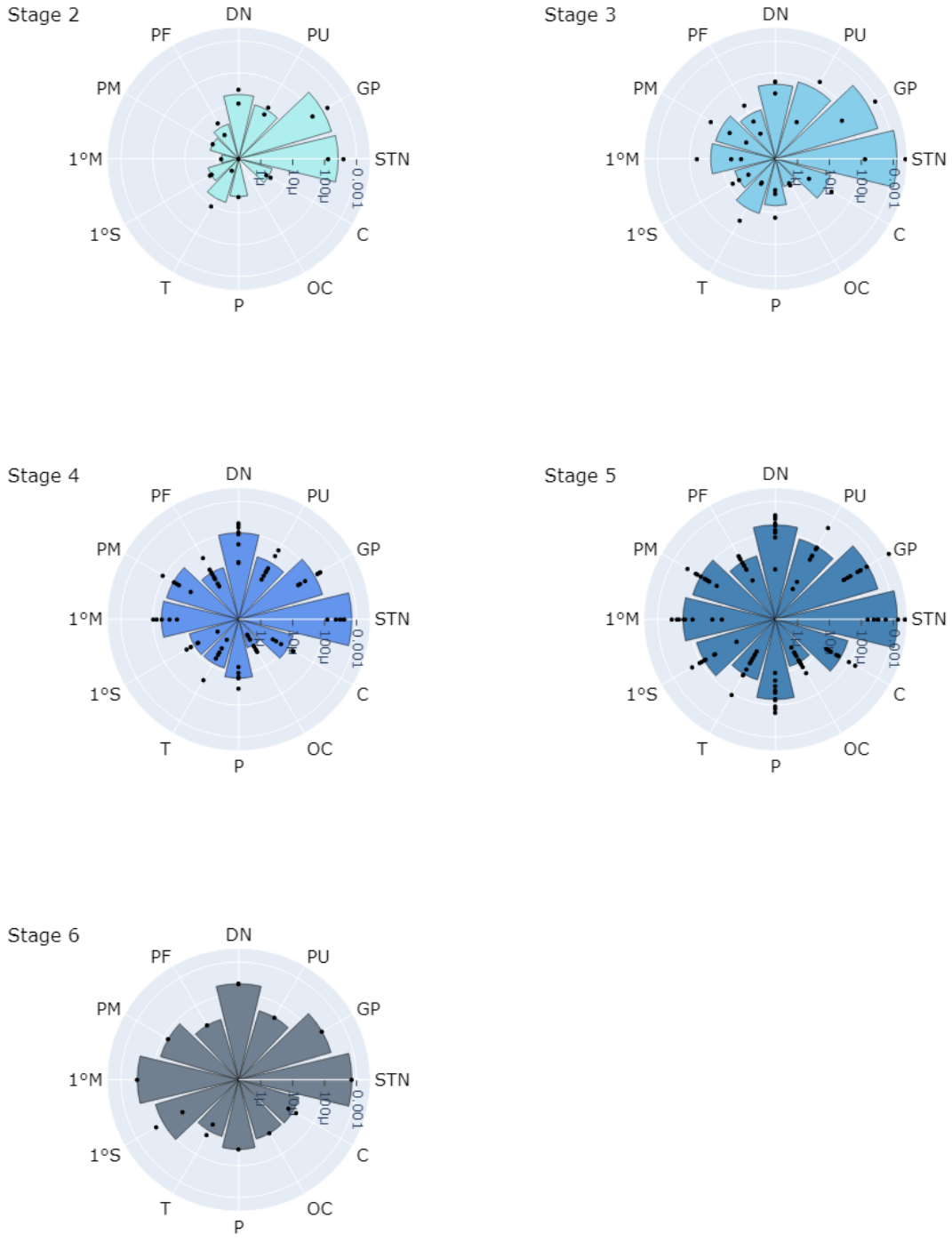



Figure 5.1: Logarithmic total tau density plot from PSP stage 2 (least severe) to 6 (most severe) across all PSP participants from both cortical and subcortical structures. STN subthalamic nucleus, GP globus pallidus, PU putamen, DN dentate nucleus, PF pre-frontal, PM pre-motor, 1°M primary motor, 1°S primary somatosensory, T temporal, P parietal, OC occipital, C cingulate.

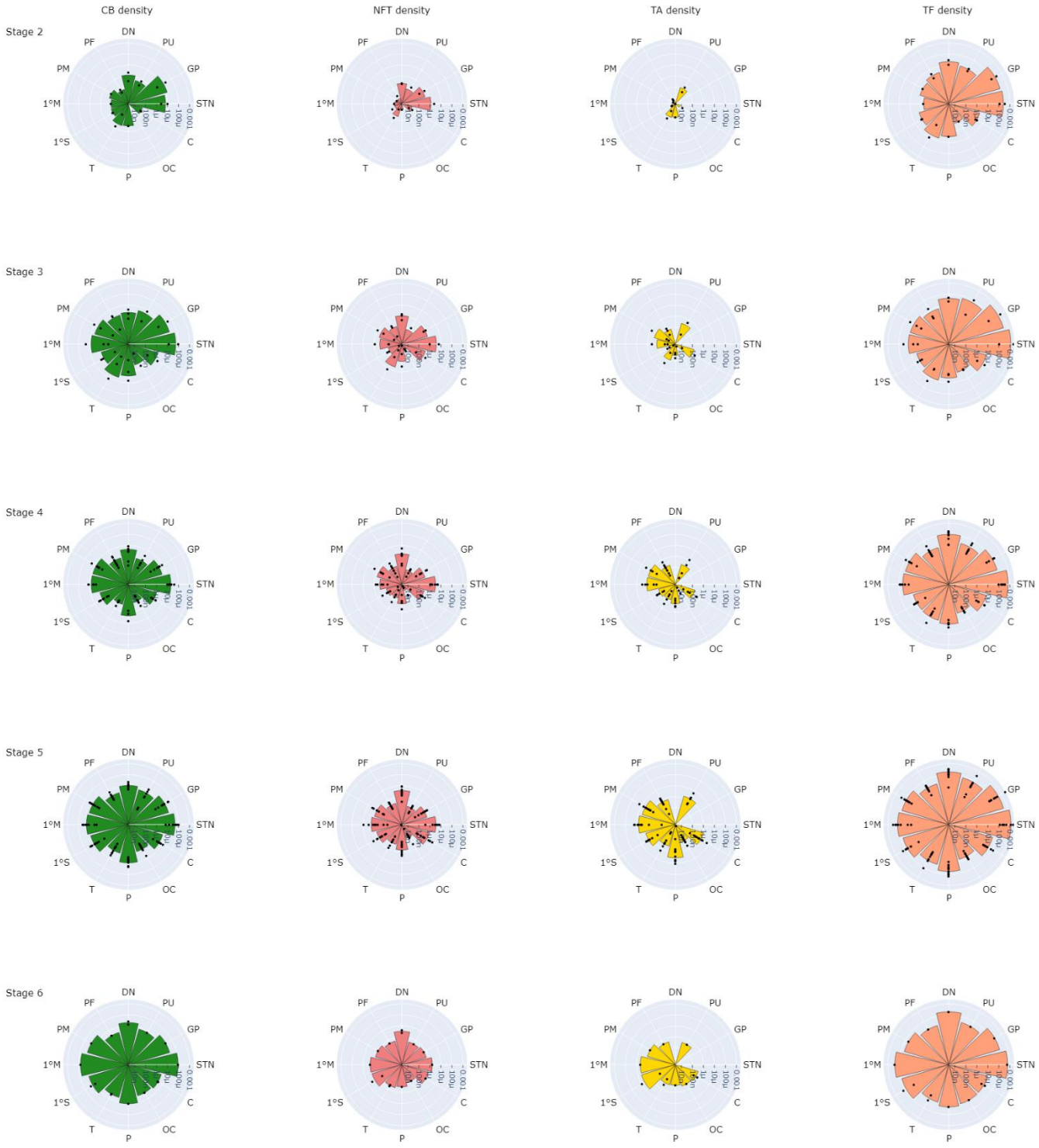



Figure 5.2: Logarithmic tau density plot by tau type per PSP stage from stage 2 (top) to 6 (bottom) across all PSP participants and brain regions sampled. CB density plot (green), NFT density plot (red), TA density plot (yellow) and TF density plot (orange). STN subthalamic nucleus, GP globus pallidus, PU putamen, DN dentate nucleus, PF pre-frontal, PM pre-motor, 1°M primary motor, 1°S primary somatosensory, T temporal, P parietal, OC occipital, C cingulate.

### 5.3.2. Correlation between tau aggregate types

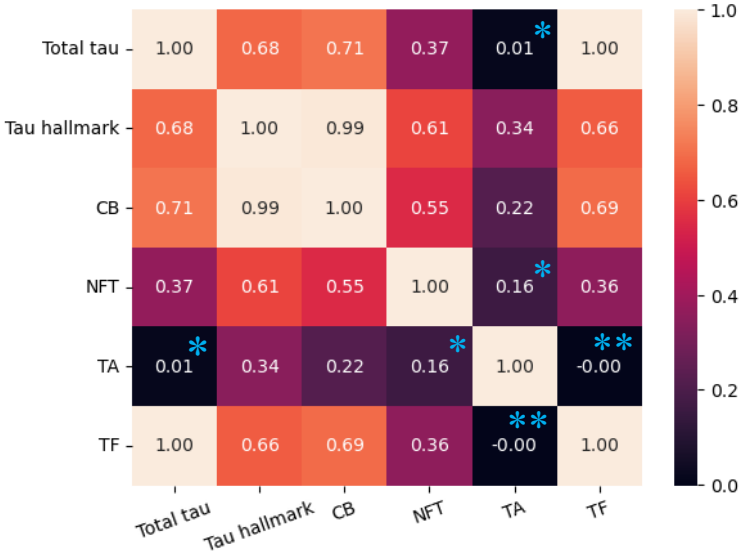
Correlation between tau aggregate types were plotted across all brain regions, cortical and subcortical regions (Figure 5.3). Across *all spatial levels*, total tau density has a perfect correlation with TF density ($r=1.00$, $p<0.001$), followed by strong correlation with CB density ($r=0.71$ to $0.89$, $p<0.001$). There was a strong correlation between total tau and NFT density across cortical regions ($r=0.69$, $p<0.001$), but modest correlation across the whole brain ($r=0.37$, $p<0.001$) and subcortical regions ($r=0.29$, $p=0.007$). Total tau had no significant correlation with TA density at the whole brain level ($r=0.01$, $p=0.84$) but strong correlation across cortical regions ($r=0.88$, $p<0.001$).

Tau hallmark density (all tau types excluding TF) generally showed strong positive correlation with total tau density ($r=0.68$ to $0.95$, $p<0.001$). CB density ($r=0.98$ to $1.00$, $p<0.001$) showed correlation to tau hallmark density across *all spatial levels*, followed by NFT density ($r=0.50$ to $0.73$, $p<0.001$). Tau hallmark density showed strong correlation with TA density in cortical regions ($r=0.78$, $p<0.001$) but not at the whole brain level ($r=0.34$, $p<0.001$).

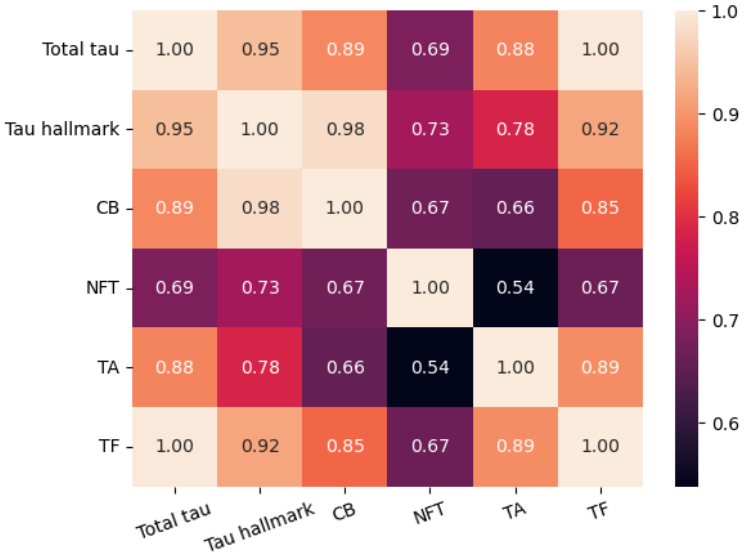
Next, the relationship between tau aggregate types were investigated. The relationship between CB or NFT density with other tau aggregate types was the consistent across *all spatial levels*, unlike for TA and TF density. CB density correlated most with TF density ($r=0.69$ to $0.85$, $p<0.001$), followed by NFT density ($r=0.44$ to $0.67$, $p<0.001$) and TA density ($r=0.22$ to $0.66$, $p<0.001$). NFT density correlated most strongly with CB density, followed by TF density ($r=0.28$ to $0.67$, $p<0.05$) and TA density ($r=0.16$ to $0.54$, $p<0.05$). TA density, at the *whole brain level*, showed a weak correlation with CB density ($r=0.22$, $p<0.001$), followed by NFT density ($r=0.16$, $p=0.008$) and no correlation with TF density ($r=0.00$, $p = 0.99$). The order changed slightly across cortical regions where TA density correlated most strongly with TF density ($r=0.89$, $p<0.001$). TF density, at the *whole brain level*, showed the highest correlation with CB density ($0.67$, $p<0.001$), followed by NFT density ($0.36$, $p<0.001$) and no correlation with TA density. Across *cortical level*, TF density correlated most strongly with TA density ($r=0.89$, $p<0.001$), followed by CB density ($r=0.85$, $p<0.001$) and NFT density ($r=0.67$, $p<0.001$). The order was preserved at subcortical level with CB density ($r=0.71$, $p<0.001$) showing higher correlation to TF density than NFT density ($r=0.28$, $p=0.007$).

Taken together, these findings demonstrate that total tau density was mainly driven by TF density and when only focusing on tau hallmark density, CB density was the main driver across all spatial levels. In general, the different type-specific tau aggregates generally correlated with one another. Specifically, CB correlated most strongly with TF density, followed by NFT density across all spatial levels and TA density correlated most strongly with other tau types particularly across cortical regions.

a) Whole brain




b) Cortical regions




c) Subcortical regions

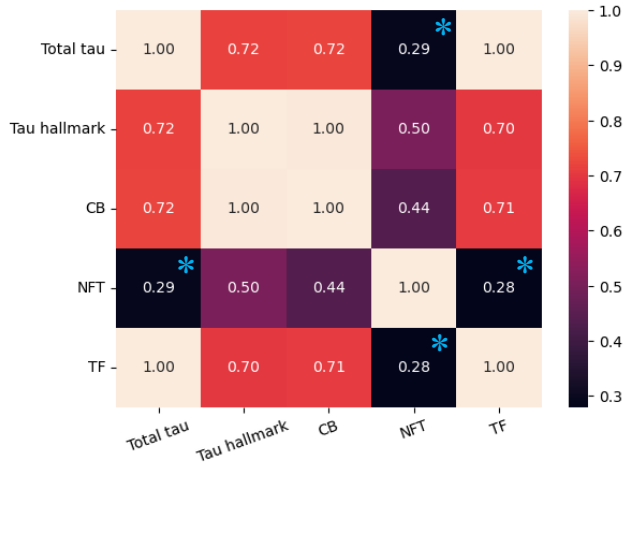



Figure 5.3: Correlation matrices of tau aggregate types across a) all regions, b) cortical and c) subcortical regions in the study. All correlation coefficients are significant at p<0.001. **0.001<p-value<0.05, **p-value > 0.05,*

### 5.3.3. Correlation to the current PSP staging scheme

Table 5.4: Spearman's correlation coefficients between tau density and PSP stage when considering all brain regions, only cortical regions, and only subcortical regions. *Correlations significant at P <0.05, ** Correlations are significant at P<0.001. Not applicable (N/A) as TA density is only quantifiable in putamen.

| Tau type | All regions | Cortical regions | Subcortical regions |
|---|---|---|---|
| Total tau density | 0.37** | 0.57** | 0.27* |
| CB+NFT+TA density | 0.47** | 0.59** | 0.39** |
| CB density | 0.46** | 0.58** | 0.38** |
| NFT density | 0.37** | 0.51** | 0.24* |
| TA density | 0.37** | 0.62** | N/A |
| TF density | 0.36** | 0.56** | 0.26* |

Across all brain regions in the study, there was a positive correlation between tau hallmark (CB+NFT+TA) density quantified from all regions and the overall PSP stage (Table 5.4). CB and TA densities showed the strongest correlation to PSP stage when considering only cortical structures. NFT density also generally showed positive correlation to PSP stage, but the correlations were weaker than that of the glial tau.

Table 5.5: Spearman's correlation coefficients between tau density from each region in the PSP staging system, and the overall PSP stage. Globus pallidus (GP), subthalamic nucleus (STN), putamen (PU), dentate nucleus (DN), pre-frontal (PF) and occipital (OC) regions. *Correlations are significant at $p<0.05$, **Correlations are significant at $p<0.001$. Not applicable (N/A) where TA density is not quantifiable.

| Tau type | GP | STN | PU | DN | PF | OC |
|---|---|---|---|---|---|---|
| Total tau density | 0.19 (p=0.41) | 0.34 (p=0.13) | 0.28 (p=0.22) | 0.63** | 0.70** | 0.85** |
| Tau hallmark density | 0.30 (p=0.19) | 0.51* | 0.37 (p=0.10) | 0.71** | 0.75** | 0.81** |
| CB density | 0.30 (p=0.19) | 0.49* | 0.37 (p=0.10) | 0.72** | 0.67** | 0.79** |
| NFT density | 0.12 (p=0.61) | 0.12 (p=0.60) | 0.53* | 0.68** | 0.54* | 0.69** |
| TA density | N/A | N/A | 0.33 (p=0.14) | N/A | 0.73** | 0.83** |
| TF density | 0.16 (p=0.48) | 0.34 (p=0.13) | 0.27 (p=0.24) | 0.60* | 0.67** | 0.84** |

Next, we investigated the contribution of tau quantified at each region in the PSP staging system (Table 5.5) to the overall PSP stage. Total tau and tau hallmark density in the occipital region showed the highest correlation to the overall PSP stage, followed by pre-frontal, dentate nucleus, subthalamic nucleus, putamen and globus pallidus respectively. These trends are in-line with the defining features of PSP staging where subcortical regions are heavily affected early in the disease stage therefore tau density in these regions is less informative in distinguishing between higher PSP stages than tau density in cortical regions that is a feature of mid to late disease stages.

When looking at individual tau type-specific densities, CB density in the globus pallidus/subthalamic nucleus and dentate nucleus showed the strongest contribution in comparison to other tau types to PSP stage. In contrast, NFT density in putamen, TF and TA density in the occipital region, and TA density in the pre-frontal region showed the strongest contribution to overall PSP stage when compared to other region-specific tau densities.

Table 5.6: Spearman's correlation coefficients between tau density from each region in the PSP staging and region-specific severity rating. Globus pallidus (GP), subthalamic nucleus (STN), putamen (PU), dentate nucleus (DN), pre-frontal (PF) and occipital (OC) regions. *Correlations are significant at $p<0.05$, **Correlations are significant at $p<0.001$. Not applicable (N/A) where TA density is not quantifiable.

| Tau type | GP | STN | PU | DN | PF | OC |
|---|---|---|---|---|---|---|
| Total tau density | 0.69* | 0.66* | 0.87** | 0.51* | 0.83** | 0.84** |
| Tau hallmark density | 0.62* | 0.44 (p=0.06) | 0.83** | 0.58* | 0.73** | 0.80** |
| CB density | 0.62* | 0.50* | 0.79** | 0.63* | 0.63* | 0.79** |
| NFT density | 0.50* | 0.20 (p=0.42) | 0.61* | 0.56* | 0.44* | 0.72** |
| TA density | N/A | N/A | 0.87** | N/A | 0.81** | 0.84** |
| TF density | 0.67* | 0.66* | 0.87** | 0.48* | 0.81** | 0.82** |

Finally, we investigated the correlation between the region-specific tau density and the manually assessed region-specific severity rating to understand which tau type is most contributory to grading the severity of each region (Table 5.6). TF density in the basal ganglia nuclei showed the strongest positive correlation to region-specific severity rating when compared to other tau types. CB density in the dentate nucleus and TA density in cortical regions showed the highest correlation to manually rated region-specific severity. The correlation between NFT density and region-specific rating was lower than that of glial density across all regions. In general, the correlation strength between total tau or tau hallmark density to region-specific severity rating is similar to the highest correlation strength between each tau type-specific density to region-specific density.

### 5.3.4. PSP stage, tau burden and PSPRS scores

We assessed whether there was a relationship between clinical severity (using the last PSPRS score prior to death) and neuropathological severity (using the PSP pathology stage at *post mortem*) using Bayesian regression analysis. Across PSP stages (Figure 5.4), there was evidence that clinical severity of stage 6 patients was higher than stage 2 patients (median PSPRS=28.44, Crl 6.71 to 48.57), while there was insufficient evidence that clinical severity differed between stage 3-5 patients versus stage 2 patients.

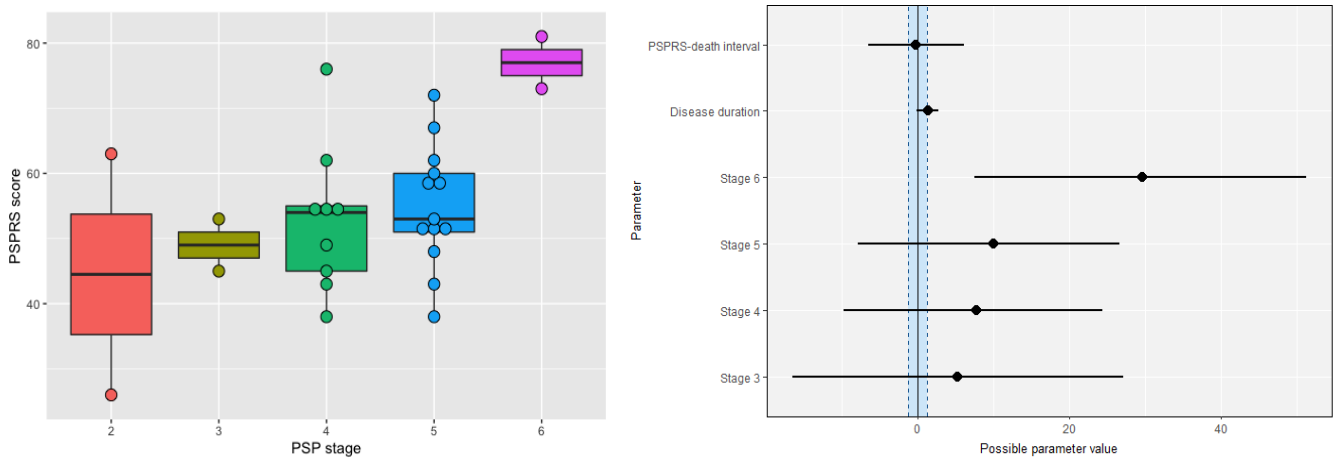



Figure 5.4: A boxplot showing PSPRS score and PSP stage (left) and a plot showing posterior distribution of the regression coefficients of the model *PSPRS score ~ PSP stage + disease duration + PSPRS-death interval* (right). Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE [-1.24 to 1.24] (blue region).

Table 5.7: Bayes' factor (BF) for each tau model compared against the total tau density model in predicting PSP rating scale (PSPRS) are presented. The comparison is made in 3 regional groupings; logarithmic tau density quantified across all brain regions, and separately for cortical and subcortical regions. **indicates BF > 3 (substantial evidence for tau type-specific density that it correlates better with PSPRS score than total tau density) or BF<1/3 (substantial evidence for total tau model as compared to tau type-specific model), and 1/3 < BF < 3 suggests the evidence from the available data is inconclusive. Not applicable (N/A) as TA density is only quantifiable in putamen.

| Model | All regions | Cortical regions | Subcortical regions |
|---|---|---|---|
| CB density | 0.69 | 0.29** | 0.47 |
| NFT density | 0.82 | 0.29** | 10.52** |
| TA density | 0.84 | 0.31** | N/A |
| TF density | 0.83 | 1.10 | 0.95 |

Next, we investigated whether total tau burden was informative of the PSPRS score when tau was quantified from all regions, only cortical and only subcortical regions. Despite a positive trend between total tau burden quantified from all regions (median = 10.96, Crl -0.24 to 21.65, 1.87% in ROPE) and only subcortical regions (median = 3.89, Crl -7.08 to 15.00, 14.79% in ROPE), there was insufficient evidence to support their relationship with PSPRS score. However, total tau burden quantified from only cortical regions (median = 10.68, Crl 2.66 to 18.91) was positively associated with PSPRS score.

To investigate whether tau type-specific burden is more informative of PSPRS score than total tau burden, we assessed tau type-specific models against a total tau model (Table 5.7). Total tau burden and tau type-specific burden were equally predictive of PSPRS score when tau was quantified from all regions. The total cortical tau burden was more predictive of PSPRS score than cortical CB, NFT and TA density but not cortical TF density (BF=1.10). However, when tau was quantified from only subcortical regions, NFT density was a better predictor of PSPRS score than total tau burden (BF = 10.52) and was chosen as the final model for subcortical tau burden. Upon final model inspection (Figure 5.5), there was decisive evidence supporting

higher tau burden and PSPRS score when tau was quantified from either only cortical or subcortical regions.

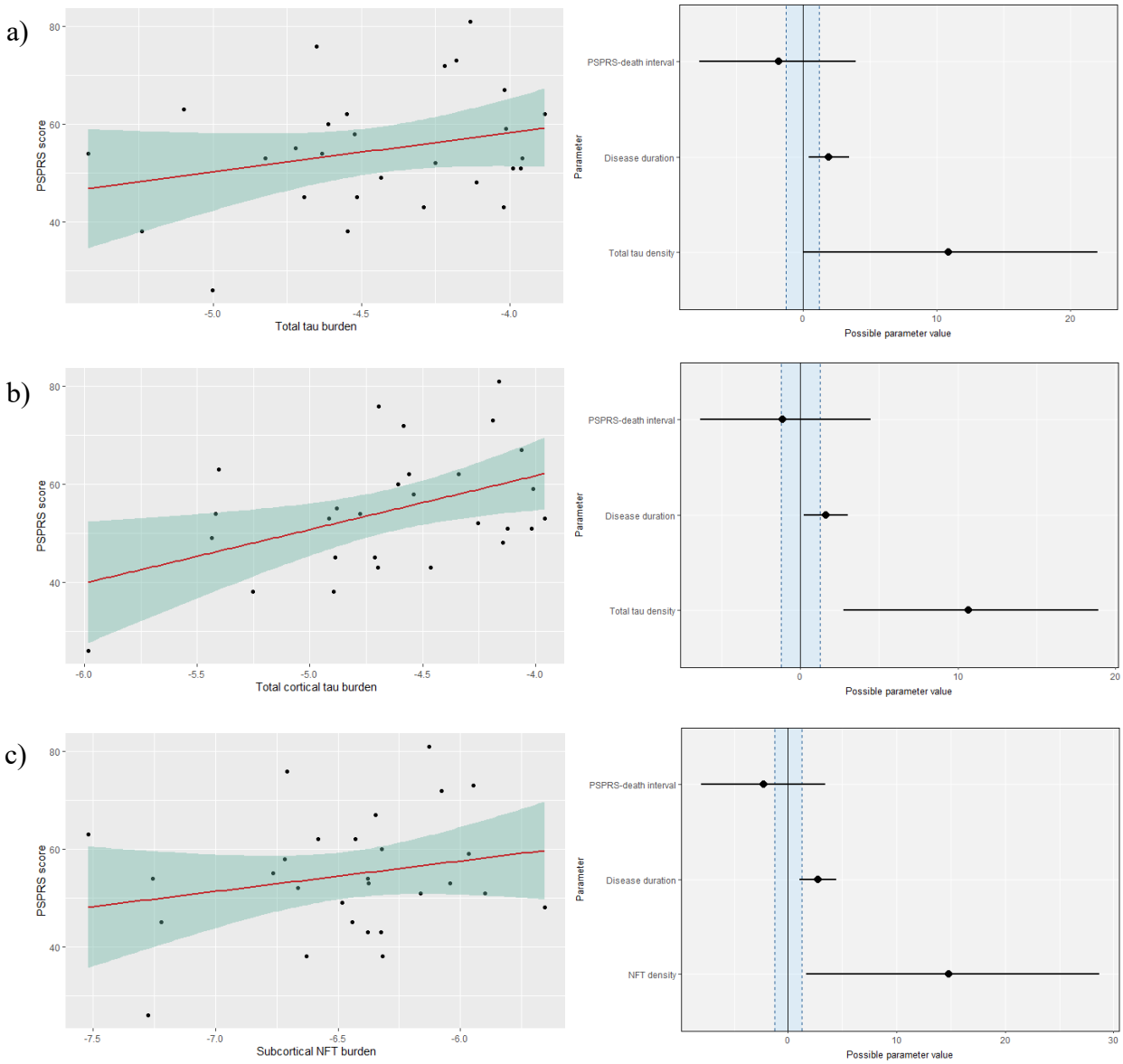



Figure 5.5: (a) A scatterplot showing PSPRS score, and total tau density quantified from all regions (left), and a plot showing the posterior distribution of the regression coefficients of the final model PSPRS score ~ tau density + disease duration + PSPRS-death interval with possible parameter values (right). Mean (circle) and 95% credible interval are plotted with ROPE [-1.24 to 1.24] (blue region). Plots from the final model when tau was quantified from only cortical regions with ROPE [-1.24 to 1.24] (b) and only subcortical regions with ROPE [-1.26 to 1.26] (c) are also presented.

### 5.3.5. Sensitivity analysis of priors

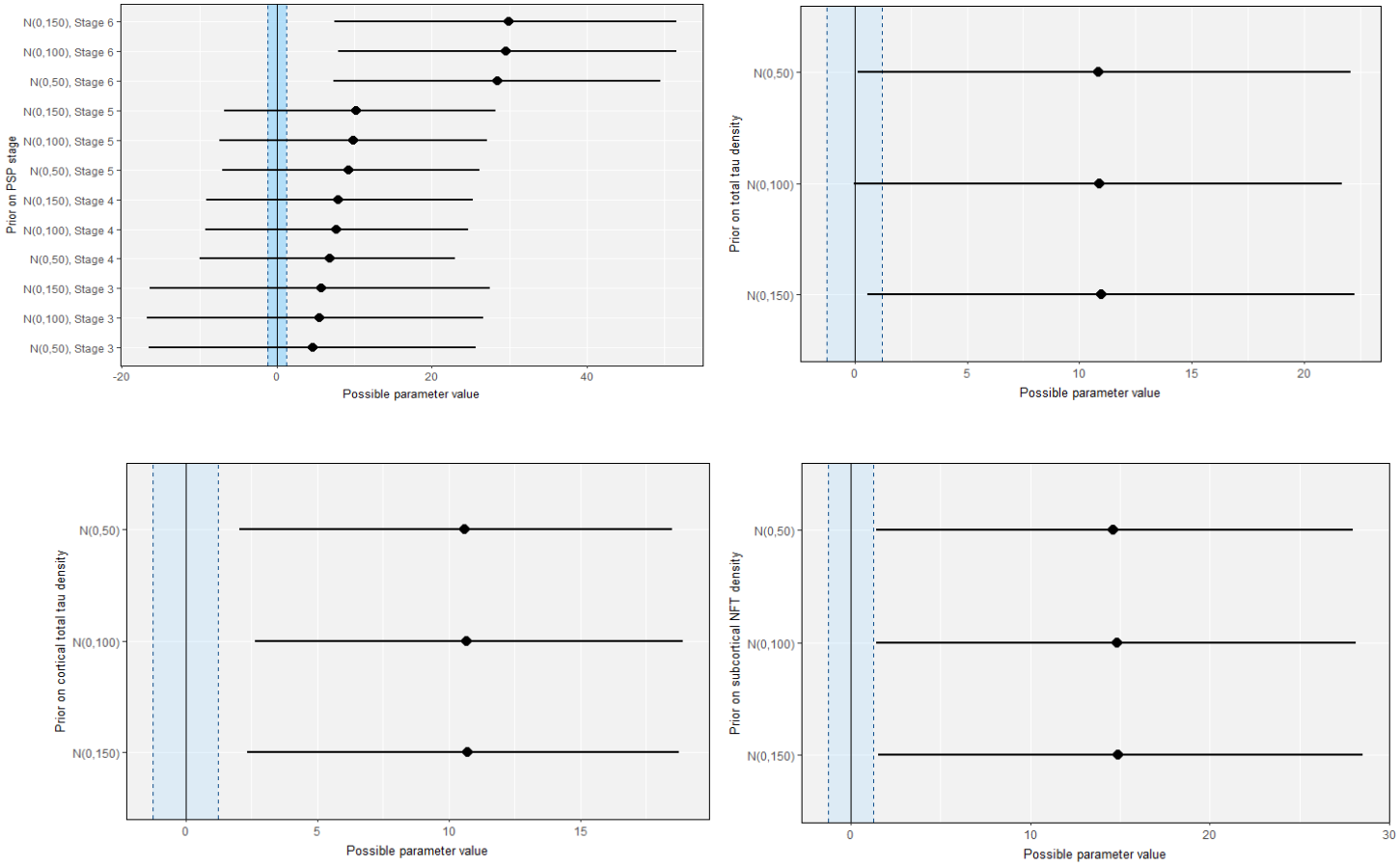



Figure 5.6: Sensitivity analysis plots showing the effect of setting alternative weakly informative priors on the regression coefficient of the effect of interest (PSP stage, tau burden) in the final models. Normal distribution, N (mean, standard deviation), was chosen with mean centred at zero, and standard deviation was varied from 50, 100 and 150.

When designing a Bayesian analysis, the choice of prior can influence the posterior distribution and the estimation of statistical differences. We assessed the sensitivity of the posterior distribution of the effects of interest (neuropathological severity) from the chosen prior choice of N (0,100) by setting other weakly informative priors. Due to the complexity of our analysis, sensitivity analysis was only conducted on the final models and posterior distributions of the neuropathological severity was qualitatively assessed. Figure 5.6 shows that choosing a less broad prior of N (0,50) or a broader prior of N (0,150) does not substantially change the conclusion of the analysis when considering ROPE: the results are robust across other weakly informative prior choices.

### 5.3.6. Total tau burden: PSP vs control subjects

We quantified total tau density in available brain slides from controls and compared to the corresponding brain regions in PSP (Figure 5.7). For controls, a small amount of tau burden could be detected, particularly in parietal, cingulate and basal ganglia nuclei. Pre-frontal, temporal and occipital regions in showed higher tau burden that other regions in controls. When compared to the corresponding brain regions from PSP participants, PSP showed higher tau burden than controls, except for the occipital region. Upon visual inspection of the occipital slides (Figure 5.8), there was one control slide with high tau burden which was likely due to age-related tau astrogliopathy (ARTAG). Other occipital slides showed minimal tau burden, in keeping with the reported tau burden from the pipeline.

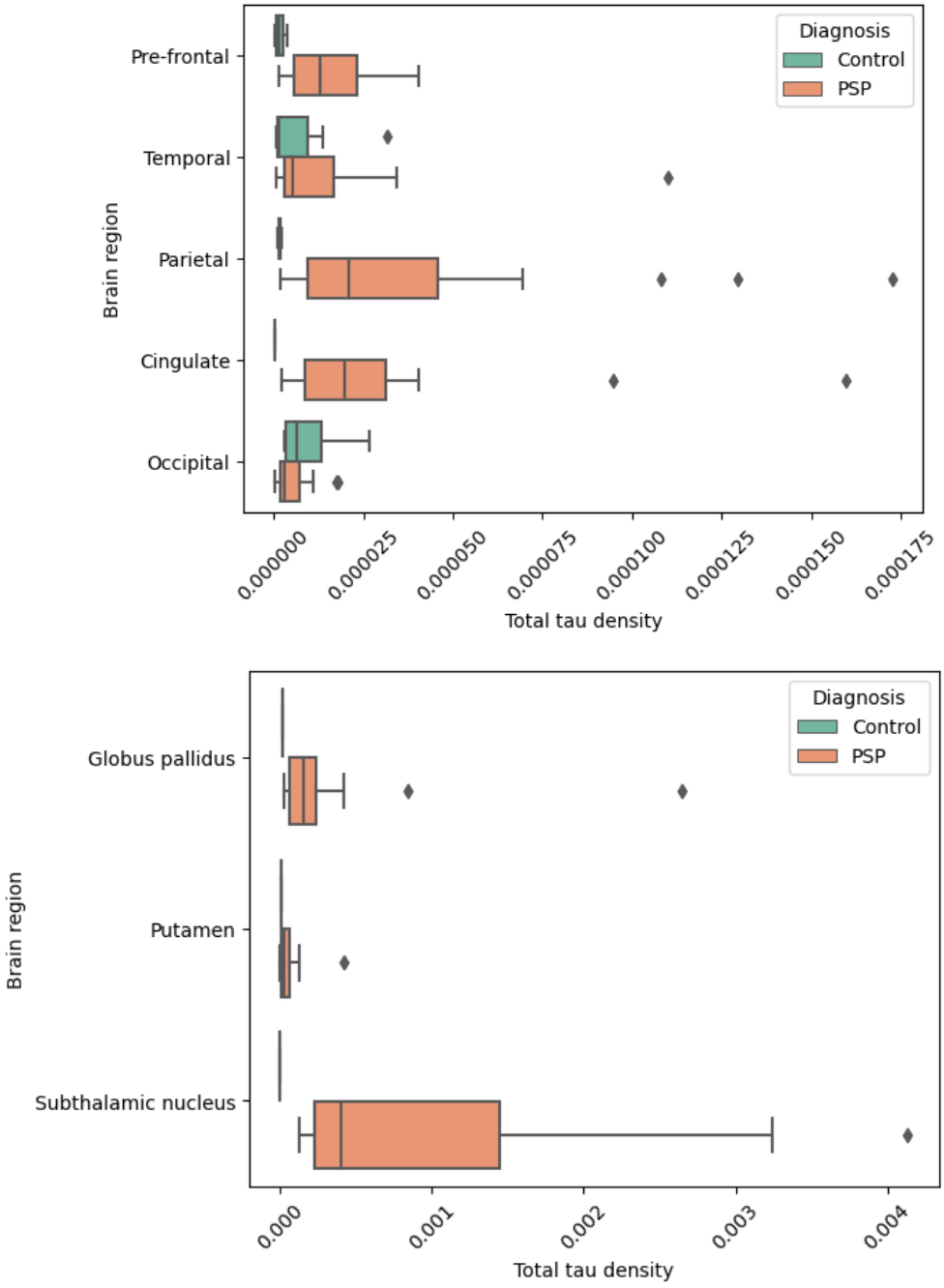
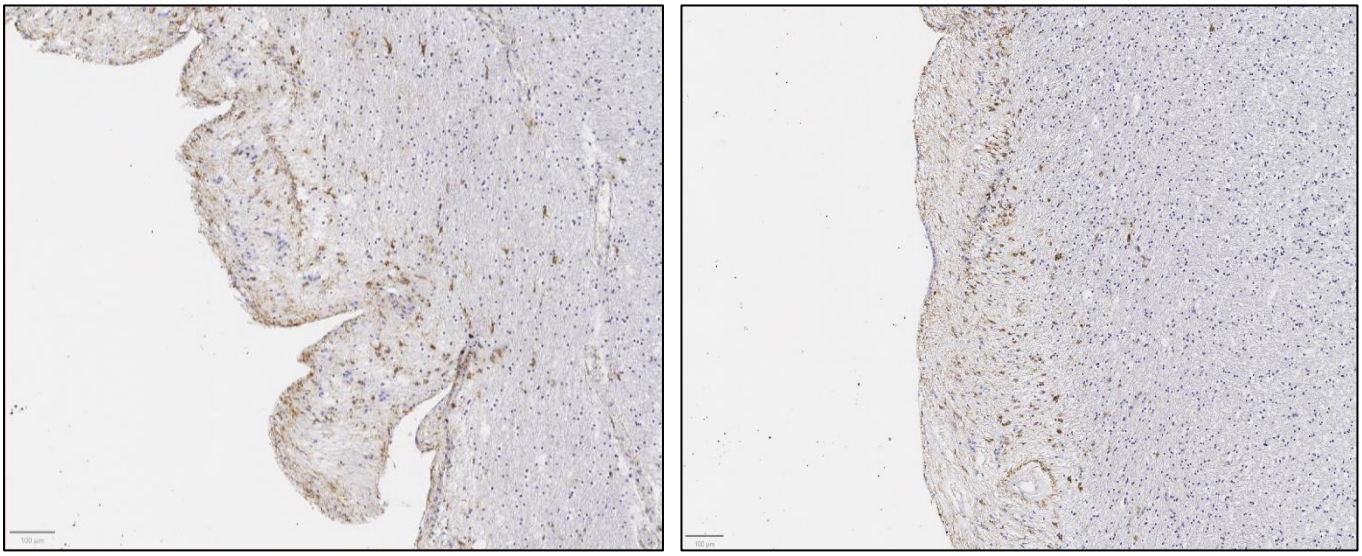



Figure 5.7: Boxplots showing total tau density quantified for each brain region between PSP and control subjects.

a)



b)

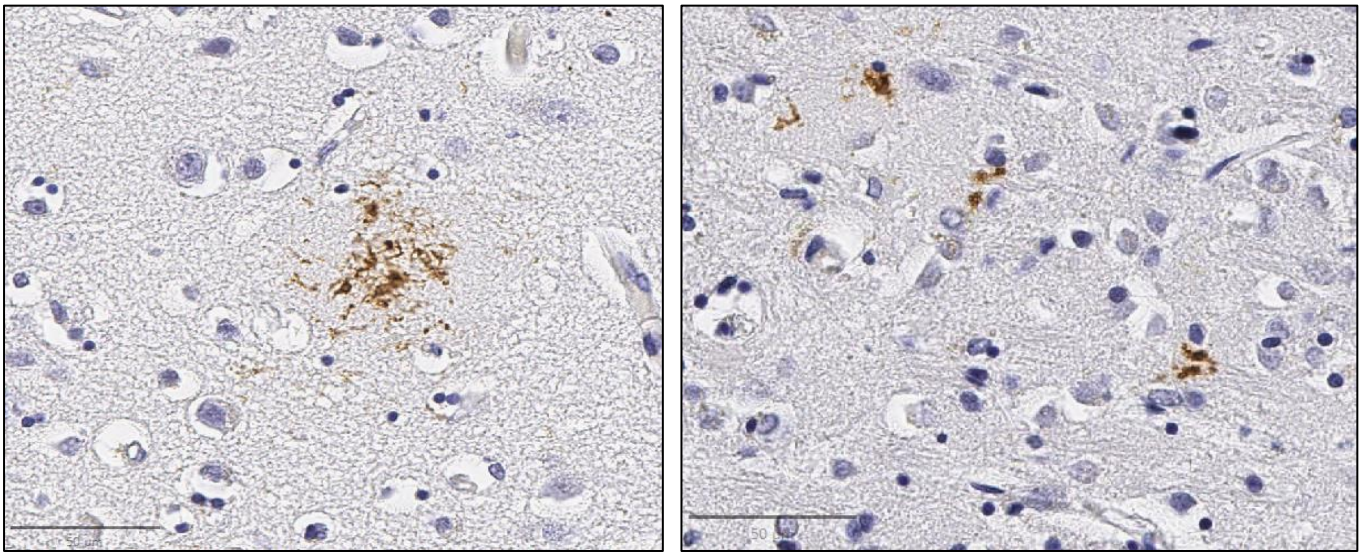


Figure 5.8: Examples from two occipital slides from control subjects. a) age-related tau astrogliopathy is observed (with scale bar 100μm) and b) tau aggregates that can generally be found in some control slides (with scale bar 50μm).

## 5.4. Discussion

### 5.4.1. Overview

By assessing all the major brain regions relevant to the established pathological staging of PSP, we were able to validate the pipeline against the best current PSP pathology staging scheme (Kovacs et al., 2020), providing confidence in the robustness of the pipeline and additional insights into PSP tau pathology. We found a strong correspondence between our automated tau quantification and the standard, manual staging approach. We were able to demonstrate that TA density in cortical areas showed the strongest relation to PSP stage, which is consistent with the PSP staging system where TA density is the focus in cortical regions when grading severity.

We went further by investigating which regions are the most informative towards PSP staging, finding that the dentate nucleus, frontal and occipital regions were more informative than basal ganglia nuclei. This is consistent with the known severe involvement of the basal ganglia nuclei from stage 2 onwards as per the described PSP staging; as a result, the severity of pathology in the basal ganglia contributes less to distinguishing between higher PSP stages. The occipital region involvement corresponds to the PSP stage 4 and onwards, which likely explains the strong correlation between tau hallmark density in the occipital lobe and PSP stage.

To build on these insights, we investigated the relationship between measured tau type density and manually assessed region-specific severity rating to understand which tau types most influence the severity rating. We found that TA (and tau fragments) density correlated strongest with the severity rating in the putamen. These findings demonstrate the utility of automated quantitative neuropathology to validate and investigate the staging and progression of tau neuropathology.

Given our algorithm's ability to quantify distinct types of tau inclusion, we investigated how the quantity and type of tau inclusions were related to clinical severity prior to death. Although there is some uncertainty given the small numbers at each stage, my results suggest that the most advanced PSP neuropathology stage 6 had the most severe clinical syndrome

measured by PSPRS scores. I demonstrated the advantage of the quantified approach by showing that that cortical tau density and subcortical NFT density were strongly associated with clinical severity measured by the PSPRS. We found largely insufficient evidence to demonstrate a linear relationship between tau burden and PSPRS score when quantified from all regions in the study. The use of Bayesian statistics indicated that more data would help to test (accept or reject) this association. I was also able to demonstrate that in subcortical regions only NFT burden is associated with disease severity. Overall, these results highlight the importance of being able to assess tau type-specific burden in specific anatomical locations, instead of using either stage alone or total tau burden in all regions.

We further validated the tau quantification pipeline on control slides and compared these to the tau burden detected on PSP slides. The result was, as expected, that control slides have less tau burden than PSP slides, suggesting that the pipeline can reliably detect tau aggregates with minimal noise. As control slides are not expected to have PSP-related tau pathology, we did not perform tau classification step on the slides. Nevertheless, the pipeline identified control slides with unusually high amount of tau, such as those with ARTAG. This could be useful in processing control slides in batch, to flag up some slides for further inspection. It also suggests that the current pipeline is potentially generalisable to other types of tau.

### 5.4.2. Strengths and limitations

I identified some limitations of the pipeline. I excluded 8 slides with significant co-pathology as the pipeline was trained to only recognised PSP-related tau pathology. The pipeline relies on DAB thresholding to detect tau objects and struggled with brain region with iron granules such as the basal ganglia. In this study, we manually removed iron granules, which was a time-consuming step that can be prone to error. Automating iron granule removal is challenging since they are heterogeneous between slides. Lastly, this study included few early stage PSP donors (stage 2 or less), who are relatively rare in brain bank cohorts (Williams et al., 2007). Nevertheless, we were able to observe the expected pattern of progression across stages from the current dataset.

Having a reliable and robust automated quantification of tau pathology will catalyse future analysis to better understand the progression of tau pathology in PSP. We anticipate our

approach can be adapted to other similar neurodegenerative tauopathies and proteinopathies. This will enable analysis of neuropathology at scale across brain regions and larger numbers of participants than is currently possible.

# Chapter 6: Neuropathological correlates of imaging measures in PSP

## 6.1. Introduction

In this chapter, I aimed to investigate the relationship between quantitative *post mortem* tau density and *in vivo* structural and functional imaging measures in PSP. This will help determine how well *in vivo* imaging measures relate to the underlying tau pathology, validating their utility as surrogate biomarkers for diagnosis and prognosis. Addressing this question will improve our understanding of the relationship between the molecular level of tau aggregation, and the macroscopic level of changes in brain structure and function.

MRI-based measures, specifically structural MRI, are widely used as part of the clinical diagnosis of patients with neurodegenerative diseases (Höglinger et al., 2017; Neary et al., 1998; Román et al., 1993). For a full evaluation of neuroimaging biomarkers, detailed validation with the underlying neuropathology is required to assess how well they reflect the underlying disease, and which aspects of neuropathology they best represent (Rittman, 2020). In AD, it is well established that atrophy on structural MRI is related to tau neurofibrils on neuropathology examination (Apostolova et al., 2015; Burton et al., 2009; Jack et al., 2002; Whitwell et al., 2012). However, atrophy may be related to different neuropathological mechanisms in distinct diseases, and little is known about the imaging-tau relationship in PSP.

One study (Spina et al., 2019) found that atrophy in PSP may be related to neuronal loss and that tau aggregation was related to atrophy across the whole brain. Their sample size was relatively small, and they employed a semi-quantitative assessment of tau burden. Their findings have yet to be validated directly, but a related study (Carlos et al., 2022) investigated atrophy-tau relationship in 4R-tauopathies (PSP and CBD combined) and found that tau relates to atrophy only in subcortical but not cortical regions. Therefore, further investigation is required to elucidate the differences in their findings.

To my knowledge, there has been only one study that investigated the relationship between functional measures and *post mortem* tau burden specifically in PSP (Spina et al., 2019). Most studies tend to use *in vivo* tau burden measured from Positron Emission Tomography (PET) (Cope et al., 2018; Franzmeier et al., 2020; Hoenig et al., 2018) and *post mortem* data of the same subject is rarely available (Lee et al., 2023). Cope et al (2018) found that nodes with stronger functional connections accrued higher tau burden than nodes that are weakly connected in AD, but the opposite association was observed in PSP. They reasoned that this could possibly explain why AD pathology is widespread in the cortex, targeting large networks and PSP pathology is concentrated in smaller number of subcortical regions. Therefore, it is of interest to see if we could replicate the findings found in PSP cohort as tau PET has off-binding target issues, especially in non-AD disorders (Lee et al., 2023; Lowe et al., 2016; Marquié et al., 2017; Sander et al., 2016). Also, using *post mortem* tau means we can investigate the relationship between functional imaging measures and tau type-specific burden, which is not possible with tau PET. Therefore, the relationship between tau burden and atrophy or functional networks require further investigation, especially with quantitative measures of tau burden.

### 6.1.1. The present study

In this study, I aimed to explore the relationship between *in vivo* structural and functional imaging measures and tau burden quantified digitally. I focused on key pathological regions in PSP subjects (Kovacs et al., 2020), covering a wide range of cortical regions, basal ganglia and dentate nucleus. For functional measures, I have chosen 2 simple graph metrics, *weighted degree* which measures the number and strength of connections of each node in the network, and *closeness centrality* which concerns how efficient the communication of each node in the network is.

Hypotheses:

1. PSP patients will show atrophy across cortical and subcortical regions, where the degree of atrophy in subcortical regions will be greater than in cortical regions.
2. PSP patients will show variations in functional measures across regions.

3. Given that previous studies found that tau is associated with atrophy, particularly in subcortical regions (Carlos et al., 2022; Spina et al., 2019), I anticipate that such a relationship will be found in subcortical regions.
4. Given the role of tau in microtubule stability, tau dysfunction and accumulation, tau burden would be related to impairments in functional network properties. Reflecting Cope et al (2018), I anticipate higher *in vivo* tau is associated with higher weighted degree and lower closeness centrality in PSP.
5. Structural and functional changes relate differentially to tau type-specific burden, as neuronal and glial cells have different roles in the brain and may respond differently during pathogenesis (Forrest et al., 2019).

This work was performed in collaboration with Timothy Rittman (Senior Clinical Research Fellow) and David Whiteside (NIHR Academic Clinical Research Fellow) who pre-processed fMRI data and extracted graph metrics.

## 6.2. Materials and methods

### 6.2.1. Subject and brain regions

A total of 154 subjects, comprising 133 healthy controls (mean age 67.98 years, range 58.90-84.50; 61 male), and 21 PSP (mean age 73.87, range 55.00-84.10; 9 male) were included in the study (Table 6.1). All PSP patients were diagnosed with either probable or possible PSP based on published diagnostic criteria (Höglinger et al., 2017) and the clinical diagnoses were verified by *post mortem* examination. All subjects had MRI scans performed at the Wolfson Brain Imaging Centre. Of 21 PSP subjects, 15 of them have fMRI data (11 multi-echo scans, 4 single-echo scans).

Table 6.1: Demographics table of patients with progressive supranuclear palsy. Not applicable (N/A) where information is missing from the database.

| Subject | Clinical diagnosis | Gender | Age at Death | MRI-to-death Interval (years) | fMRI-to-death Interval (years) | Disease duration (years) |
|---|---|---|---|---|---|---|
| 1 | poss. PSP-CBS | Female | 78.9 | 0.81 | 0.81 | 5 |
| 2 | poss. PSP-CBS | Female | 79.8 | 1.81 | 1.81 | 3.42 |
| 3 | poss. PSP-PGF | Male | 75.2 | 0.41 | 1.37 | 4.51 |
| 4 | poss. PSP-SL | Female | 78.4 | 2.42 | 2.42 | 8.83 |
| 5 | prob. PSP-F | Male | 75.4 | 1.12 | 4.99 | 8.42 |
| 6 | prob. PSP-RS | Male | 63 | 2.3 | 2.3 | 8.83 |
| 7 | prob. PSP-RS | Female | 71.6 | 0.26 | 0.26 | 4.58 |
| 8 | prob. PSP-RS | Female | 78.1 | 1.76 | N/A | 16.75 |
| 9 | prob. PSP-RS | Female | 76.4 | 1.73 | 1.73 | 8.75 |
| 10 | prob. PSP-RS | Male | 80.5 | 0.93 | N/A | 6.5 |
| 11 | prob. PSP-RS | Female | 69.9 | 0.4 | 0.4 | 5.33 |
| 12 | prob. PSP-RS | Male | 71 | 2.68 | N/A | 5.42 |
| 13 | prob. PSP-RS | Male | 74.5 | 2.22 | N/A | 6 |
| 14 | prob. PSP-RS | Male | 77.4 | 1.95 | N/A | 6.33 |
| 15 | prob. PSP-RS | Female | 84.1 | 0.62 | N/A | 4.25 |
| 16 | prob. PSP-RS | Female | 65.3 | 1.06 | 1.06 | 13.92 |
| 17 | prob. PSP-RS | Female | 71.9 | 0.79 | 3.54 | 6.17 |
| 18 | prob. PSP-RS | Female | 55 | 2.82 | 2.82 | 5.5 |
| 19 | prob. PSP-RS | Female | 74.7 | 3.08 | 3.08 | 6.58 |
| 20 | prob. PSP-RS | Male | 71.5 | 1.77 | 1.77 | 5.17 |
| 21 | prob. PSP-RS | Male | 78.7 | 0.59 | 2.32 | 5.33 |

For clinical diagnosis of PSP participants, *prob.* probable, *poss.* Possible, *RS* Richardson syndrome, *CBS* predominant corticobasal syndrome, *SL* predominant speech and language disorder, *F* predominant frontal presentation, *PGF* progressive gait freezing.

### 6.2.2. Quantitative pathology in PSP

A total of 163 slides (199 regions) were obtained (2-12 slides per brain, median =11, IQR=4). The slides included 126 neocortical slides (19 pre-frontal, 18 occipital, 15 temporal, 15 parietal, 15 cingulate, 15 primary somatosensory, 15 primary motor, 14 pre-motor slides), 18 basal ganglia slides (with globus pallidus, subthalamic nucleus and putamen) and 19 dentate nucleus slides. The tau quantification pipeline was applied to all slides to quantify tau type-specific pathologies ('coiled body' (CB), 'neurofibrillary tangles' (NFT), 'tufted astrocyte' (TA), 'tau fragments (TF)). Total tau density was quantified by summing up all tau types and divided by the area sampled. Tau hallmark density includes CB, NFT, TA but not TF.

Subcortical regions, except for putamen, have no TA quantification. See slide preparation outlined in core methods chapter.

### 6.2.3. MRI pre-processing

T1-weighted MRI scans were pre-processed using FreeSurfer, an automated surface-based analysis which parcellated cortical structures and segmented sub-cortical structures (Fischl & Dale, 2000). First, skull, eyes and neck were removed from the scans. Subcortical structures were segmented, and intensity normalization applied to enhance the distinction between grey and white matter. White matter segmentation followed, and cortical surfaces were registered to Desikan-Killiany atlas for gyral labelling. All pre-processed brains were quality checked using VisualQC (Raamana, 2018) and manually corrected using Freeview where necessary (Fischl & Dale, 2000).

Pre-processed brains were parcellated using the Brainnetome Atlas, a validated connectivity-based parcellation atlas suitable for multi-modal analyses (Fan et al., 2016). The Brainnetome atlas contains 210 cortical and 36 subcortical regions but does not contain brainstem or cerebellar regions, therefore these regions were added from the FreeSurfer segmentation. Since the subthalamic nucleus and dentate nucleus were not part of the default FreeSurfer segmentation, probability maps and masks were used to segment them from T1 images using SPM12 in MATLAB (MATLAB, 2010). A total of 258 regions from both hemispheres were used in the study.

Single-subject w-score maps of grey matter atrophy were computed, where region-specific volumes were standardized with estimated total intracranial volume prior to w-score calculation. W-scores are analogous to z-scores but have been corrected for age at scan and gender (see Figure 6.1), representing the patient's deviation from the norm, i.e. the control group (Jack et al., 1997; La Joie et al., 2012; Spina et al., 2019). Similar to z-scores, w-scores in the control group have a mean value of 0 and standard deviation of 1, where a value of +1.65 corresponds to the $95^{th}$ and a value of -1.65 corresponds to the $5^{th}$ percentile (La Joie et al., 2012). W-score maps (see Figure 6.2) were computed and reversed so that positive w-scores represent atrophy (La Joie et al., 2012).

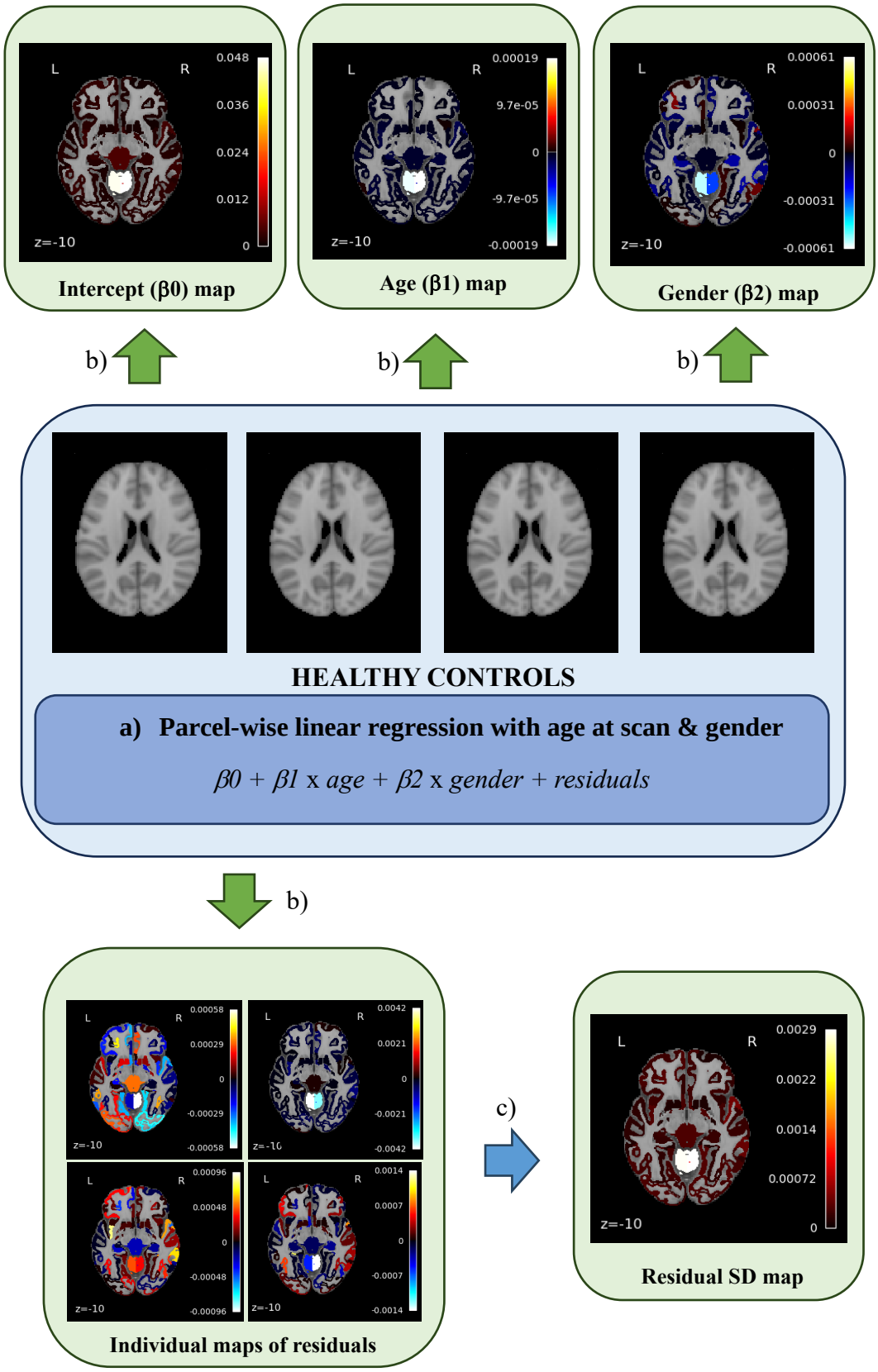



Figure 6.1: Illustration of regression coefficient map creation from control group as part of w-score calculation. a) Linear regression with covariates (age at scan and gender) was performed in the control group to estimate age and gender related changes. b) Maps with parcel-wise intercept values (β0), age-related (β1), gender-related (β2), individual maps of residuals can be extracted. c) The standard deviation (SD) of the residuals was then computed.

**W-score calculation**

In conceptual term,

$$w - score = \frac{raw\ value - value\ predicted\ for\ patient's\ age\ \&\ gender}{SD\ of\ residual\ in\ healthy\ controls}$$

In mathematical term,

$$output = \frac{p\_m\ -\ (b0\_m\ +\ (b1\_m\ x\ age)\ +\ (b2\_m\ x\ gender)\ )}{sd\_m}$$

Where p_m: patient's map, b0_m: β0 (intercept), b1_m: β1 map (age), b2_m: β2 map (gender), sd_m: residuals SD map

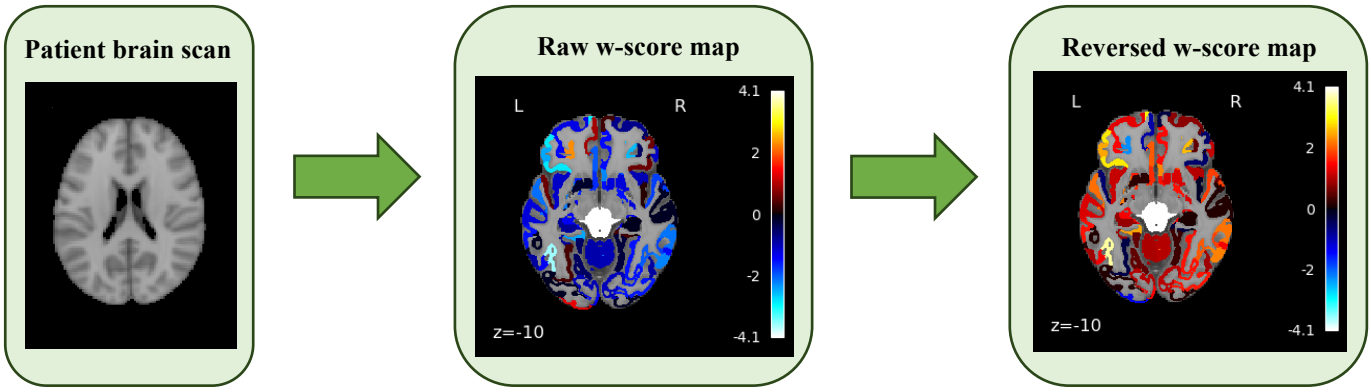



Figure 6.2: Illustration of w-score calculation after extracting maps of regression coefficient from controls. Raw parcel-wise values are extracted from each patient brain scan and w-score is calculated using the formula. Raw w-scores are reversed so that higher positive value means more atrophy.

### 6.2.4. fMRI pre-processing

Pre-processing of fMRI data is the same as a previous study from our lab (Whiteside et al., 2021). Participants underwent fMRI imaging at 3T using echo-planar imaging sensitive to the blood-oxygen-level-dependent signal (TR 2s, TE 30ms, whole brain acquisition, 3 x 3 x 3.75mm voxels, 305 volumes) with eyes open in a dark bore. High resolution T1-weighed Magnetization Prepared Rapid Gradient Echo (MPRAGE) structural images (TR 2s, TE 2.93ms, flip angle 8, voxel size 1.1mm isotropic) were acquired during the same session for use in normalization.

Image pre-processing used FSL's FEAT for registration to the structural image, motion correction, 100Hz high-pass temporal filtering, 5mm FWHM spatial smoothing. Denoising was performed using FSL's FIX with a training set of 10 subjects from each disease group per cohort. Additional removal of motion artefact used wavelet de-spiking (brainwavelet.org).

Graph theoretical analysis was performed using Maybrain software (https://github.com/RittmanResearch/maybrain) and NetworkX (Hagberg et al., 2008), with the Brainnetome parcellation. *Weighted degree* was extracted which measures the number and strength of nodal functional connections and *closeness centrality*, the inverse of the path length between a node and all other nodes in the graph. A network density threshold of 5% was used for the primary statistical analysis. I also report the results at density threshold of 1-10%, with 1% increments. A total of 244 regions were included in the analysis as graph metrics are not available for the subthalamic nucleus, cerebellum, and brainstem.

### 6.2.5. Matching imaging and pathology modalities

Table 6.2: A matching between histology slide and Brainnetome atlas parcel for all brain regions in the imaging-pathology analysis.

| Region | Histology slide | Brainnetome parcel | Region |
|---|---|---|---|
| Pre-frontal | BA46 | A9/46d_L, A46_L, A9/46v_L | Middle frontal gyrus |
| Pre-motor | BA6 | A6dl_L, A6m_L, A6vl_L, A6cdl_L, A6cvl_L | Superior frontal gyrus, precentral gyrus |
| Primary motor | BA4 | A4hf_L, A4ul_L, A4t_L, A4tl_L | Precentral gyrus |
| Primary somatosensory | BA1-3 | A1/2/3ulhf_L, A1/2/3tonla_L, A2_L, A1/2/3tru_L | Postcentral gyrus |
| Temporal | BA21/22 or BA 41/42 | A41/42_L, A22c_L, A22r_L, A21c_L, A21r_L | Superior temporal gyrus, middle temporal gyrus |
| Parietal | BA7 | A7r_L, A7c_L, A7pc_L, A7ip_L | Superior parietal lobule |
| Occipital | BA17/18 | rCunG_L, cCunG_L, cLinG_L, rLinG_L | MedioVentral occipital cortex |
| Cingulate | BA23, 24, 28-33 | A23d_L, A24rv_L, A32p_L, A23v_L, A24cd_L, A23c_L, A32sg_L | Cingulate gyrus |
| Subthalamic nucleus | Extracted from basal ganglia | Freesurfer | Subthalamic nucleus |
| Globus pallidus | Extracted from basal ganglia | GP_L | Globus pallidus |
| Putamen | Extracted from basal ganglia | vmPu_L, dlPu_L | Putamen |
| Dentate nucleus | Extracted from cerebellum | Freesurfer | Dentate nucleus |

To match pathology slides to the Brainnetome atlas, we selected Brainnetome parcels that correspond to the specific brain region in which the pathology slides were sectioned from (Table 6.2, see Appendix for parcel code details). As multiple Brainnetome parcels make up a given region, we took a mean value of those parcels for both w-score and nodal measures.

### 6.2.6. Imaging-tau analyses

For this analysis, a logarithmic transformation ($log_{10}$) was applied to tau density scores due to the skewness of the tau density distribution. To investigate the relationship between *in vivo* imaging measures and tau density, the *brms* package in R (Bürkner, 2021) was used to construct Bayesian linear mixed regression models. This analysis was carried out separately with *in vivo* structural and functional measures as the outcome variables, tau density as the fixed effect, imaging-death interval, gender, age at death, disease duration, echo acquisition (only for functional metrics) as covariates. Brain region and patient identifier were included as random effects. The relationship between imaging measures and tau density was first assessed at a *whole-brain level* to investigate if a common relationship exists across all regions. If no relationship is found, we would investigate such relationship across *cortical regions* only and within each brain region.

First, to investigate whether tau density is predictive of imaging measures, total tau and tau type-specific models were constructed. For each model, I first determined which covariates are useful by creating a full model with all covariates and assessing regression coefficients using the region of practical equivalence (ROPE). If 95% of the credible interval (Crl) of the regression coefficient falls completely within ROPE, then the effect of the parameter would be equivalent to the null value for practical purposes (Kruschke, 2011; Kruschke & Liddell, 2018a) and would be excluded from the model. Next, I assessed the total tau regression coefficient using ROPE to establish a baseline relationship. With all tau models, I then tested whether tau type-specific density was more informative of the *in vivo* imaging measures than the simpler measure of total tau density by comparing tau type-specific models to the total tau model. For the final model, the strength of regression coefficients for tau and each covariate were assessed using a ROPE.

A Gaussian model family was selected based on the distribution of the data. A relatively informed normal prior *(mean=0, SD =1)* was chosen for the tau coefficient when assessing its relationship with atrophy from the previously found effect size (Spina et al., 2019) and a weakly informative prior *(mean=0, SD =10)* was chosen when assessing tau relationship with functional measures. For analyses, weakly informative priors (*mean=0, SD=10*) were chosen for covariates, and default R priors were used for the intercept *(student-t prior; df=3, mean =*

*1, SD = 2.5* for all regions and region-specific analysis, *student-t prior; df=3, mean = 0.6, SD = 2.5* for only cortical regions*)* and the sigma *(student-t prior; df=3, mean =0, scale = 2.5).*

Due to the complexity of our analysis, a sensitivity analysis of priors was conducted to assess the effect of the prior choice on the regression coefficients in the final models. We chose two other sets of normal priors, one more informative *(*for tau-atrophy, *mean = 0, SD = 0.*5 for tau*; mean = 0, SD = 5* for covariates*;* for tau-functional measure*, mean = 0, SD = 5 for* both tau and covariates*)* and the other less informative *(*for tau-atrophy, *mean = 0, SD = 2* for tau*; mean = 0, SD = 20* for covariates; for tau- functional measures, *mean =0, SD = 20* for both tau and covariates*)* to assess the sensitivity of posterior estimates on the prior choice.

## 6.3. Results

### 6.3.1. Structural and functional measures assessment across all regions in PSP

To assess the relationship between structural and functional brain measures, atrophy and functional graph metrics calculated from PSP subjects were first visualized (Figure 6.3, for full results see Appendix). Regions with particularly high w-scores, in the 95th percentile, include subthalamic nucleus, putamen, thalamus, cerebellar white matter, midbrain and pons. Subcortical regions generally had higher w-scores than cortical regions. Frontal regions had higher w-scores than more posterior cortical regions.

Regions with particularly high weighted degree included the lateral pre-frontal thalamus and a wide range of cortical regions such as the superior frontal, pre- and para-central gyrus, superior and middle temporal gyrus, precuneus, post-central gyrus, cingulate gyrus, and rostral cuneus gyrus. Those with low weighted degree, in the bottom 5th percentile, included the caudate nucleus, and temporal regions such as the inferior temporal gyrus and parahippocampal gyrus.

Regions have very similar closeness centrality values, ranging from 0.67 to 0.75 (mean= 0.71). However, regions with high closeness centrality include amygdala, and a wide range of cortical regions such as the superior frontal, and temporal gyrus, parahippocampal gyrus, insula gyrus and cingulate gyrus. In contrast, medioventral and lateral occipital cortex have lower closeness centrality than other regions.

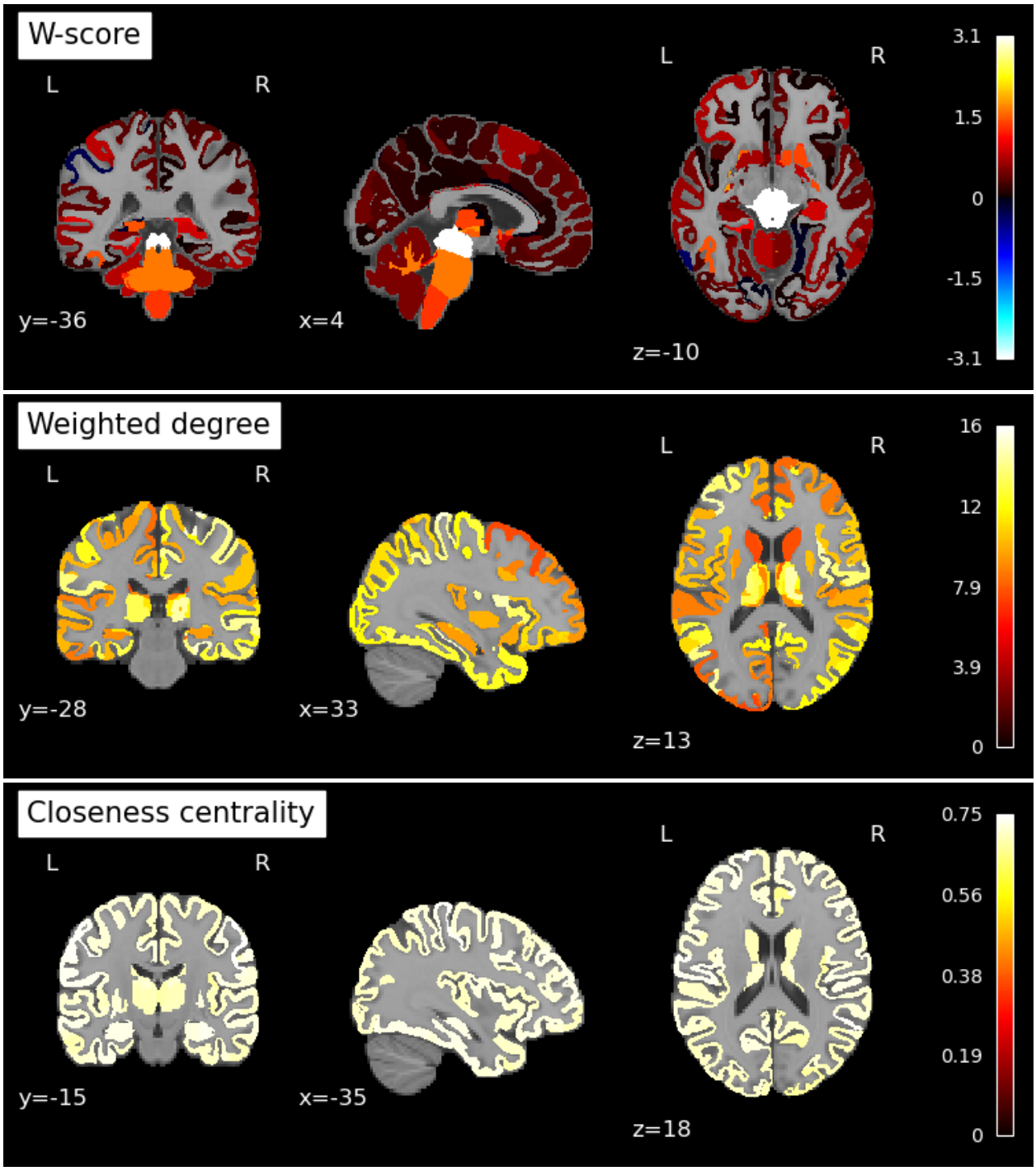



Figure 6.3: Group level w-score map of atrophy and raw functional connectivity measures of weighted degree and closeness centrality. Higher w-score corresponds to higher volume loss.

### 6.3.2. PSP w-score and tau density

Atrophy (w-scores) and tau density of the matched regions in PSP patients were firstly plotted for inspection (Figure 6.4). For atrophy, w-scores were highest in subcortical structures, particularly in the subthalamic nucleus. Cortical regions generally showed slightly lower w-scores where frontal regions showed higher w-scores than other cortical regions, and mean w-scores were lowest in parietal, cingulate and occipital regions.

Similarly, total tau density was generally higher in subcortical regions, followed by frontal regions, and was lowest in occipital cortex. For tau type-specific density, TF showed the highest density across brain regions compared to other tau types, followed by CB density, TA density (where quantifiable) and NFT density. All tau type-specific densities followed the same trend as total tau density such that subcortical regions, specifically the subthalamic nucleus showed the highest tau density, followed by frontal regions, and intermediate regions and finally the occipital region which showed the lowest tau density.

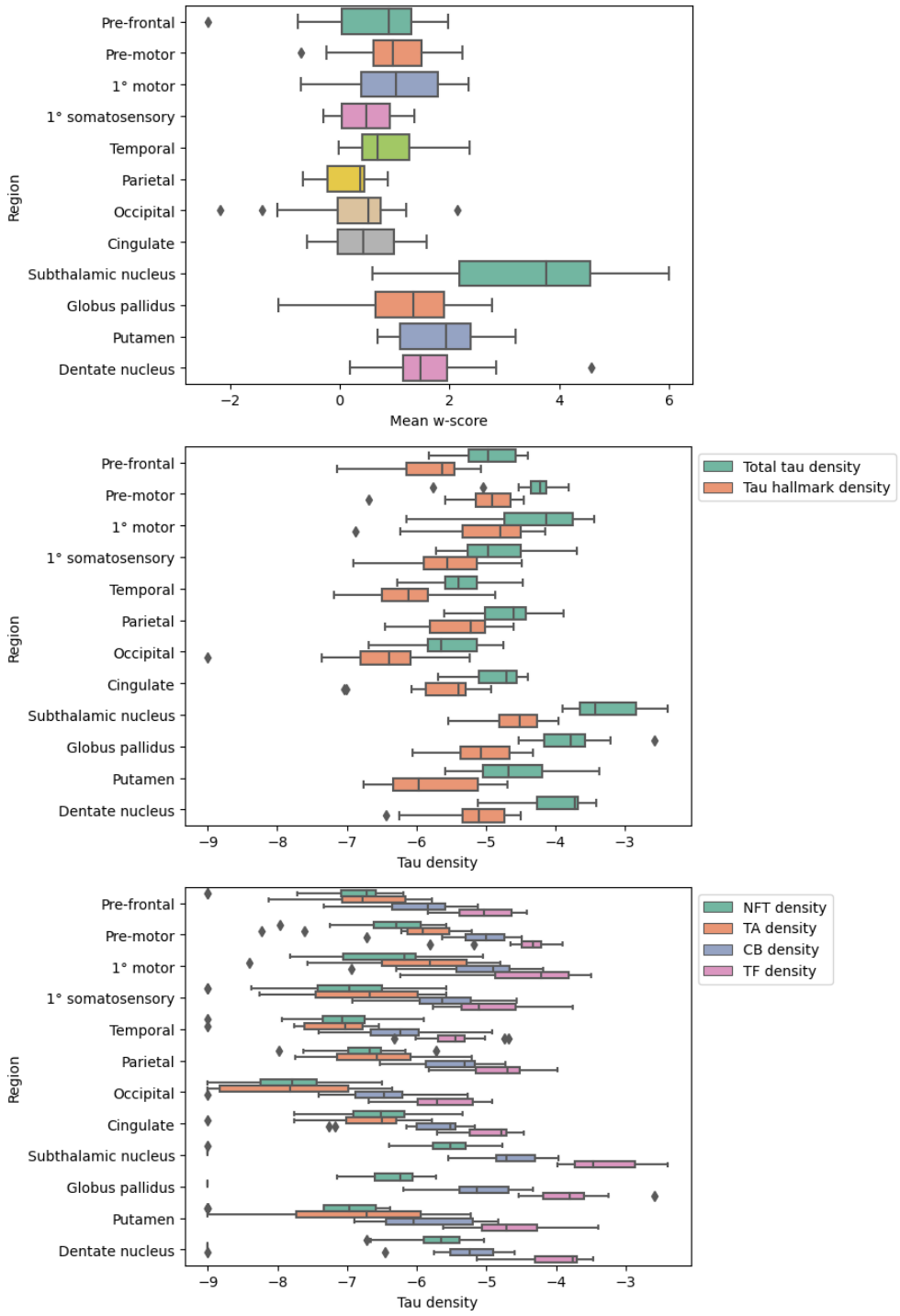



Figure 6.4: Boxplots showing a) mean w-score, b) total tau and tau hallmark density (all tau types except for TF) and c) tau type-specific density in PSP subjects across all regions with histology data. NFT neurofibrillary tangle, TA tufted astrocyte, CB coiled bodies, TF tau fragments.

### 6.3.3. Postmortem tau density is not predictive of *in vivo* whole-brain and cortical structural atrophy

Table 6.3: Bayes' factor (BF) for each tau type-specific model compared against the total tau density model in predicting w-score (atrophy) are across all and only cortical regions are presented. 1/3< BF <3 represents inconclusive evidence from model comparison.

| Model | All regions | Cortical regions only |
|---|---|---|
| Tau hallmark density | 0.57 | 0.92 |
| NFT density | 0.50 | 1.12 |
| TA density | 1.50 | 0.73 |
| CB density | 0.43 | 0.98 |
| TF density | 1.00 | 0.97 |

First, I assessed the relationship between MRI measured brain atrophy and neuropathologically quantified tau across *all brain regions* by constructing models of density for different tau types, with imaging-death interval and gender as covariates (Figure 6.5). Despite a negative trend between total tau burden and atrophy, there was insufficient evidence to support this association (median = -0.15, CrI -0.45 to 0.13, 42.00% in ROPE). To investigate whether tau type-specific burden is more informative of atrophy than total tau burden, we assessed tau type-specific models against a total tau model, but the evidence was inconclusive (Table 6.3). All tau density types showed a negative trend with atrophy, but there was insufficient statistical evidence to support this association, except for NFT density which showed a positive trend with atrophy across all regions (Figure 6.5). Sensitivity analysis of the total tau model showed robust results; the median of the posterior distribution of the regression coefficients stayed roughly unchanged for different prior choices (Figure 6.6).

Next, we assessed the relationship between MRI measured atrophy and neuropathology quantified tau across *cortical regions* in the same manner. Despite a trend for a positive relationship between total tau density and atrophy in cortical regions (Figure 6.7), there was insufficient evidence to support such relationship (median = 0.02, CrI -0.26 to 0.29, 47.63% in ROPE). There was insufficient evidence in favor of tau type-specific models as compared

to the total tau model, with the highest, but still very weak (BF = 1.12) evidence in favor of the NFT model (Table 6.3). Unlike results from the whole-brain analysis, all cortical tau type-specific densities showed a positive trend with cortical atrophy. Nevertheless, there was insufficient evidence to support these relationships. Sensitivity analysis of the total tau model showed robust estimates with other choices of priors (Figure 6.8).

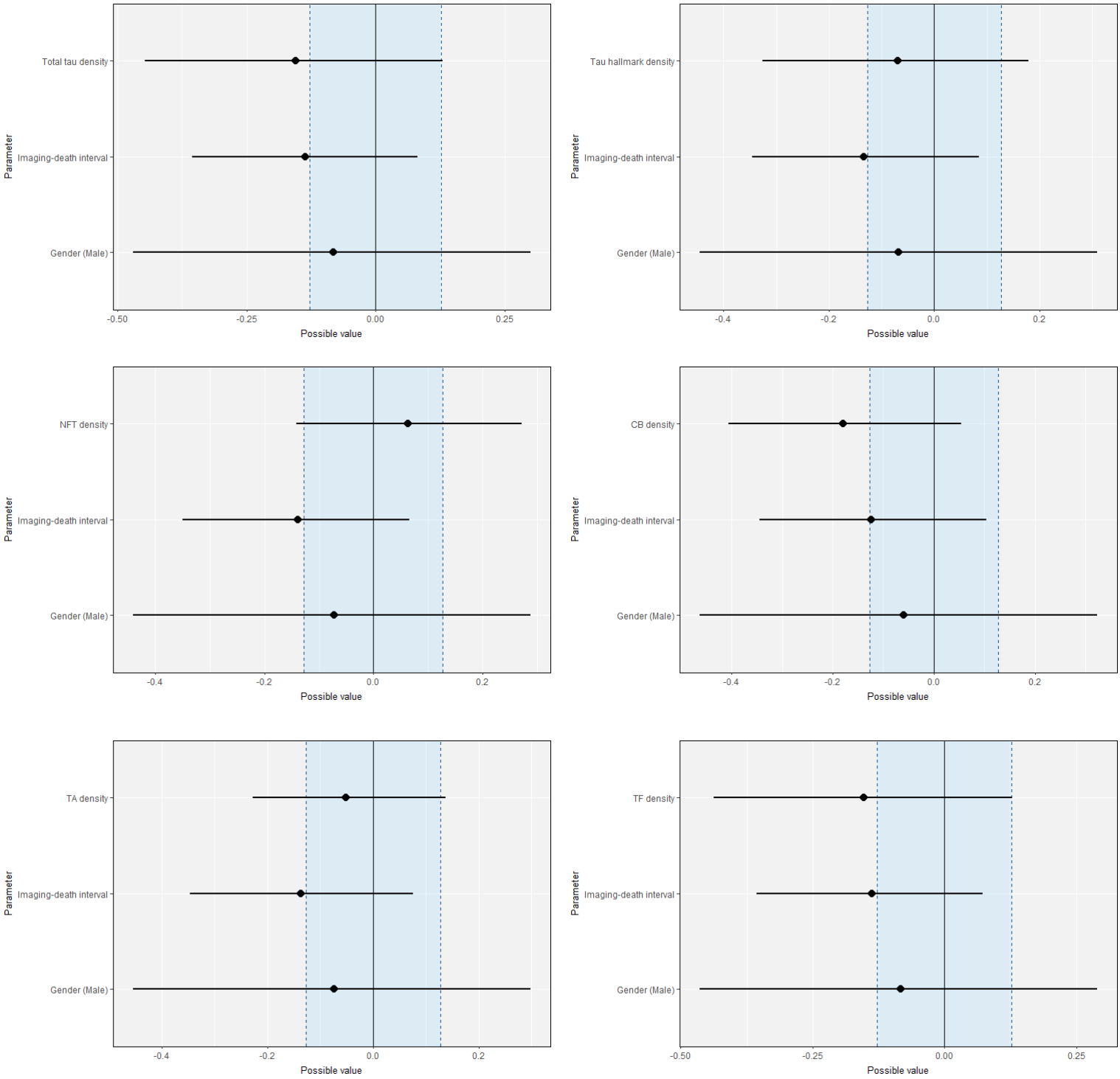



Figure 6.5: Plots showing posterior distribution of the regression coefficients of all tau models, w-score ~ tau density + imaging-to-death interval + gender + (1|region name) + (1|Patient ID) for all regions. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE [-0.13 to 0.13] (blue region).

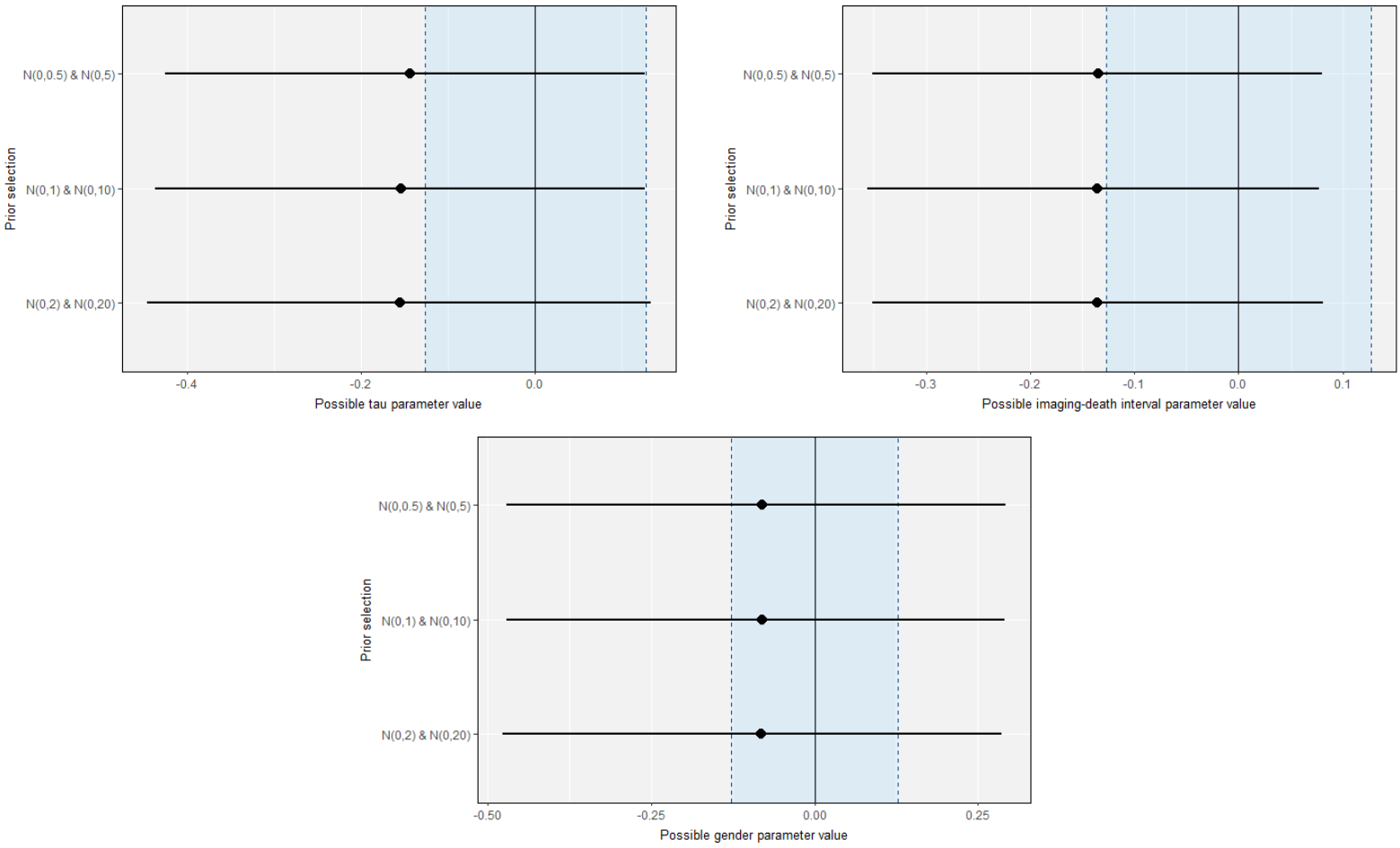



Figure 6.6: Sensitivity analysis plots of the effect of setting alternative weakly informative priors for the regression coefficient of the effect of interest, in the total tau model in predicting atrophy across all regions. A normal distribution, N (mean, standard deviation), was centered at zero and the standard deviation was varied.

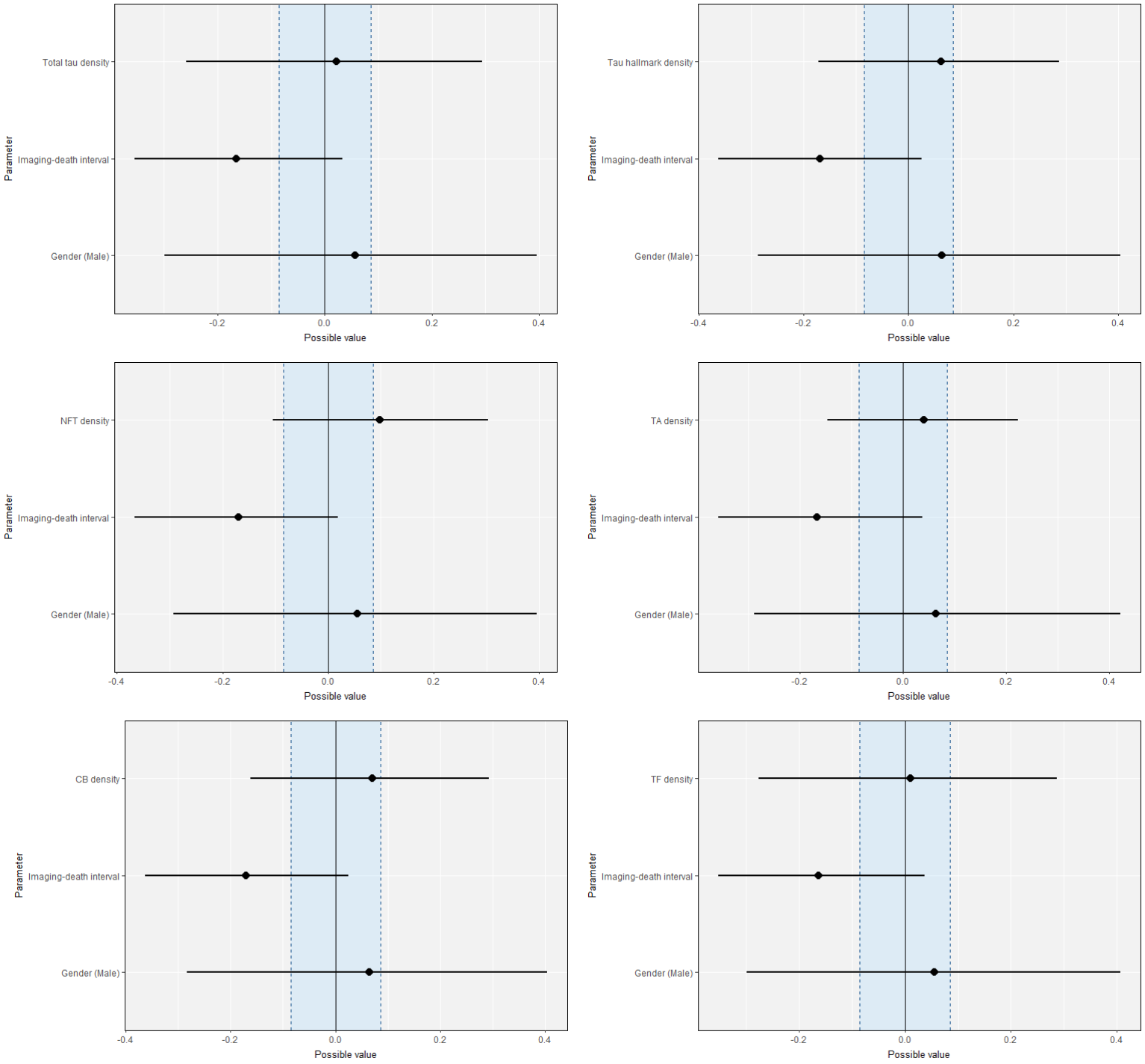



Figure 6.7: Plots showing posterior distribution of the regression coefficients of all tau models, w-score ~ tau density + imaging-to-death interval + gender + (1|region name) + (1|Patient ID) for cortical regions. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE [-0.09 to 0.09] (blue region).

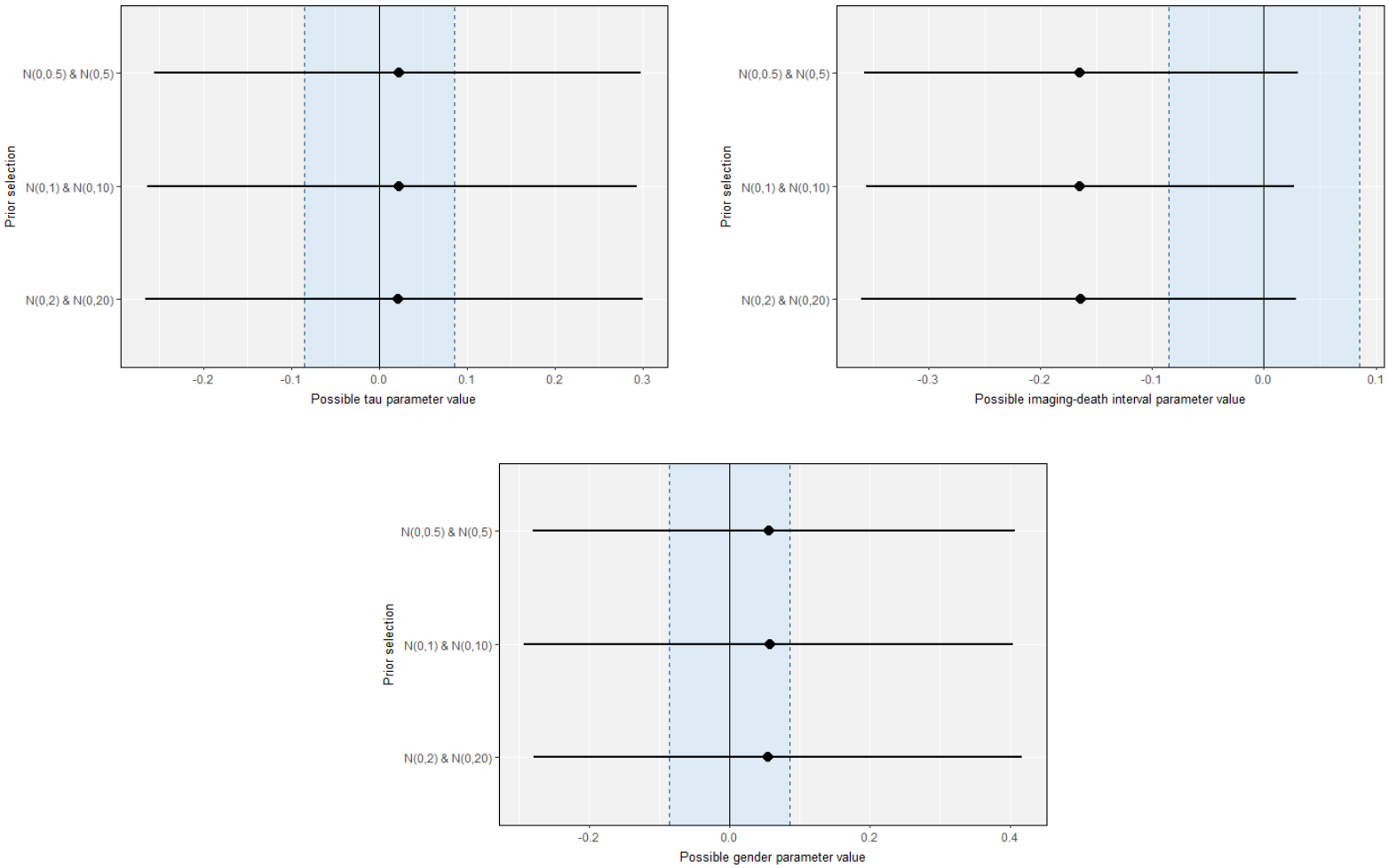



Figure 6.8: Sensitivity analysis plots for the effect of setting alternative weakly informative priors on the regression coefficient of the effect of interest in the total tau model in predicting atrophy across cortical regions. A normal distribution, N (mean, standard deviation), was centered at zero and standard deviation was varied.

### 6.3.4. Postmortem tau density may be predictive of *in vivo* dentate nucleus atrophy

To better understand the results from whole-brain and cortical analyses, we investigated the atrophy-tau relationship within each brain region across all tau types. In general, Figure 6.9 shows that there is insufficient evidence to support an atrophy-tau relationship, except for a negative association between CB density and atrophy in the dentate nucleus (median = -0.55, CrI -0.95 to -0.14, 0% in ROPE). Inspection of this negative relationship showed that it could have been driven by an outlier (Figure 6.10). To test this, the potential outlier was removed, and the analysis was re-run where there was insufficient evidence to support the relationship between atrophy and CB density in dentate nucleus (median = 0.18, CrI -0.52 to 0.90,

13.31% in ROPE). Sensitivity analysis showed that the evidence supporting atrophy-tau relationship in dentate nucleus could become inconclusive if a set of priors with narrower distributions was chosen with the full dataset, and the evidence remained inconclusive with the potential outlier removed (Figure 6.11).

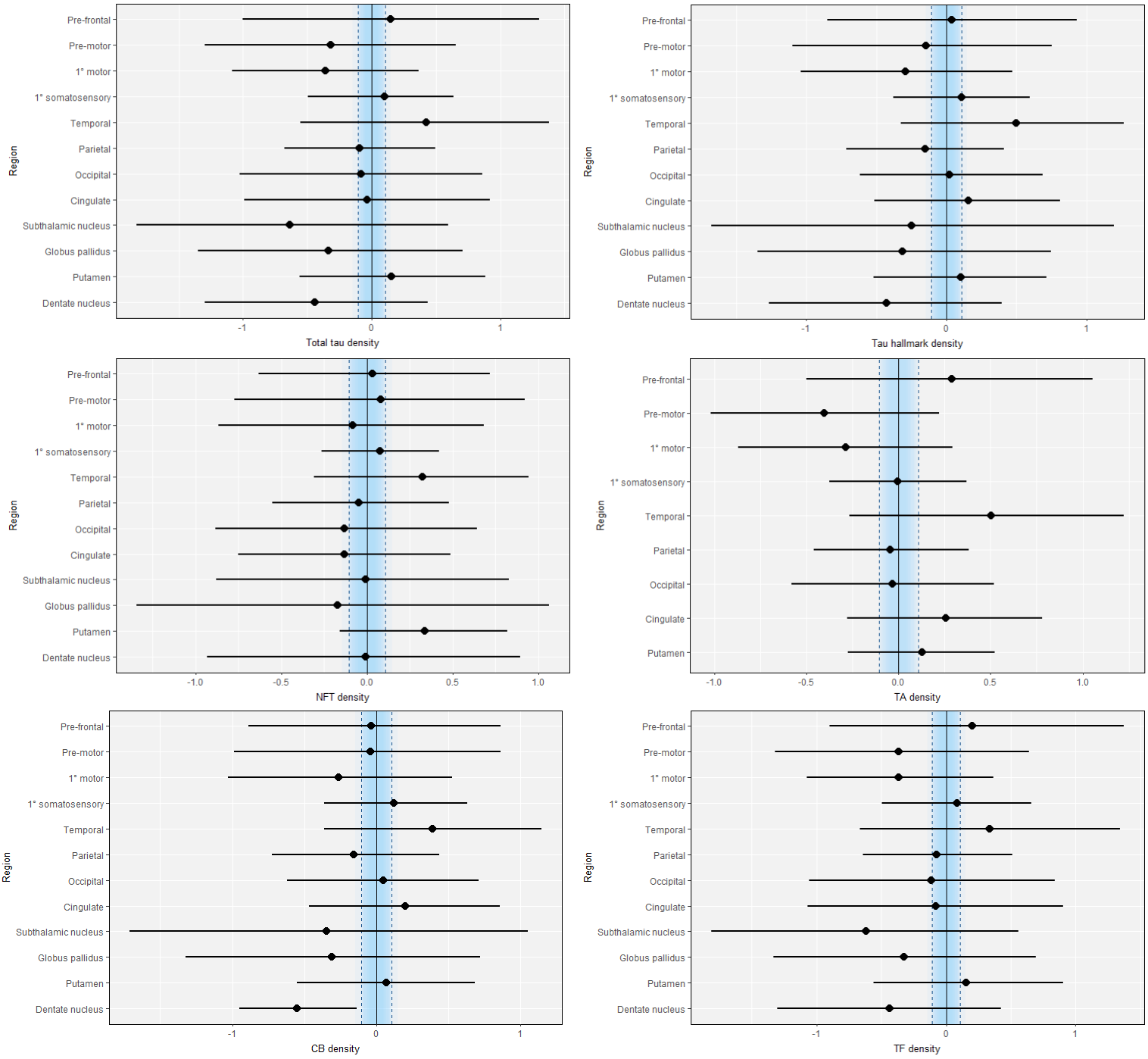



Figure 6.9: Plots showing posterior distribution of the regression coefficients of all tau models constructed for each region, w-score ~ tau density + imaging-to-death interval + gender. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE [-0.10 to 0.10] (blue region).

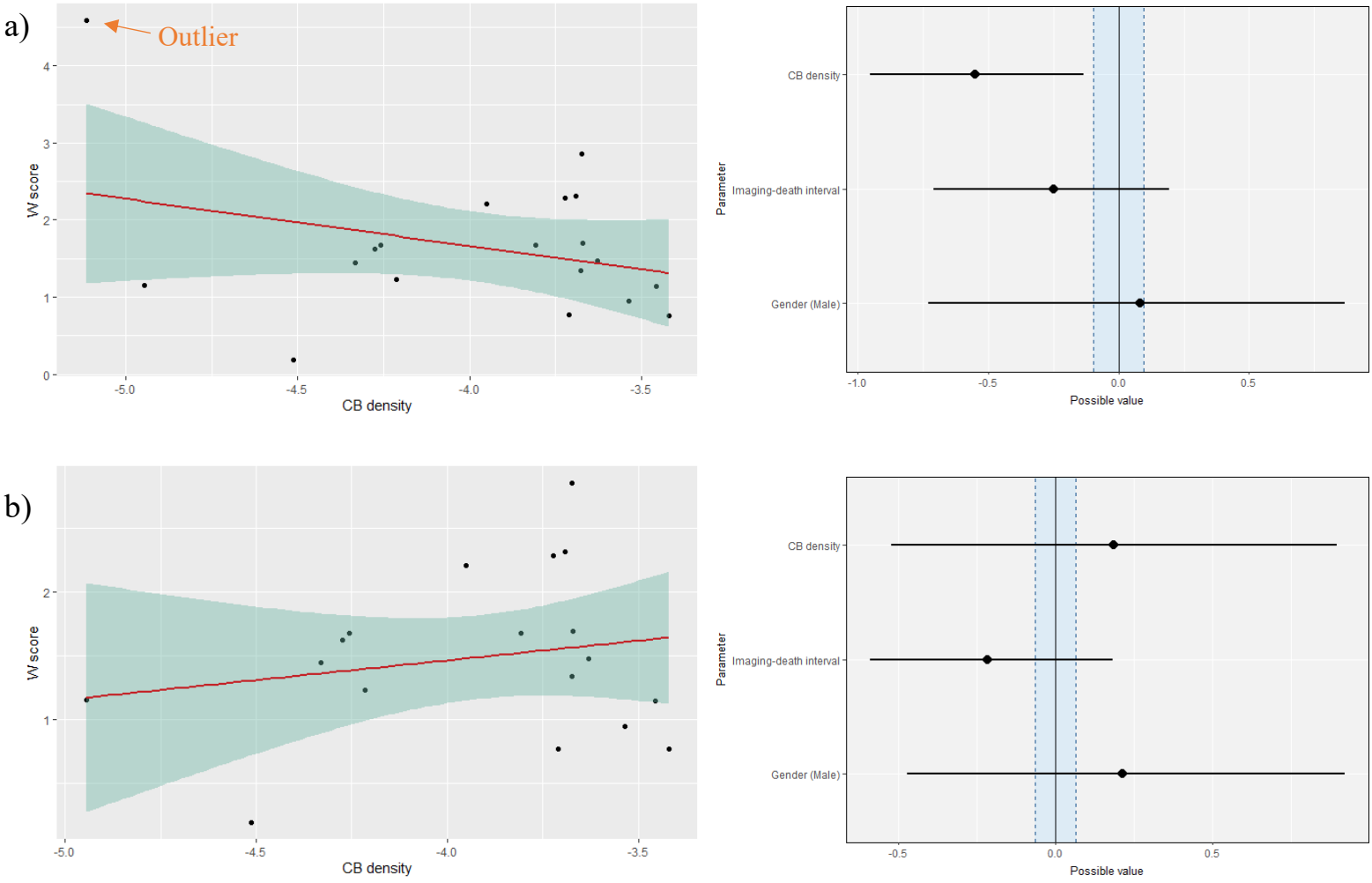



Figure 6.10: Scatterplots (left) showing CB density and w-score (atrophy) in dentate nucleus with a) full dataset and b) after removing a potential outlier. Plots (right) showing the posterior distribution of the regression coefficients of the tau models; w-score ~ CB density + imaging-death interval + gender. Mean (circle) and 95% credible interval are plotted with ROPE [-0.10 to 0.10] (blue region).

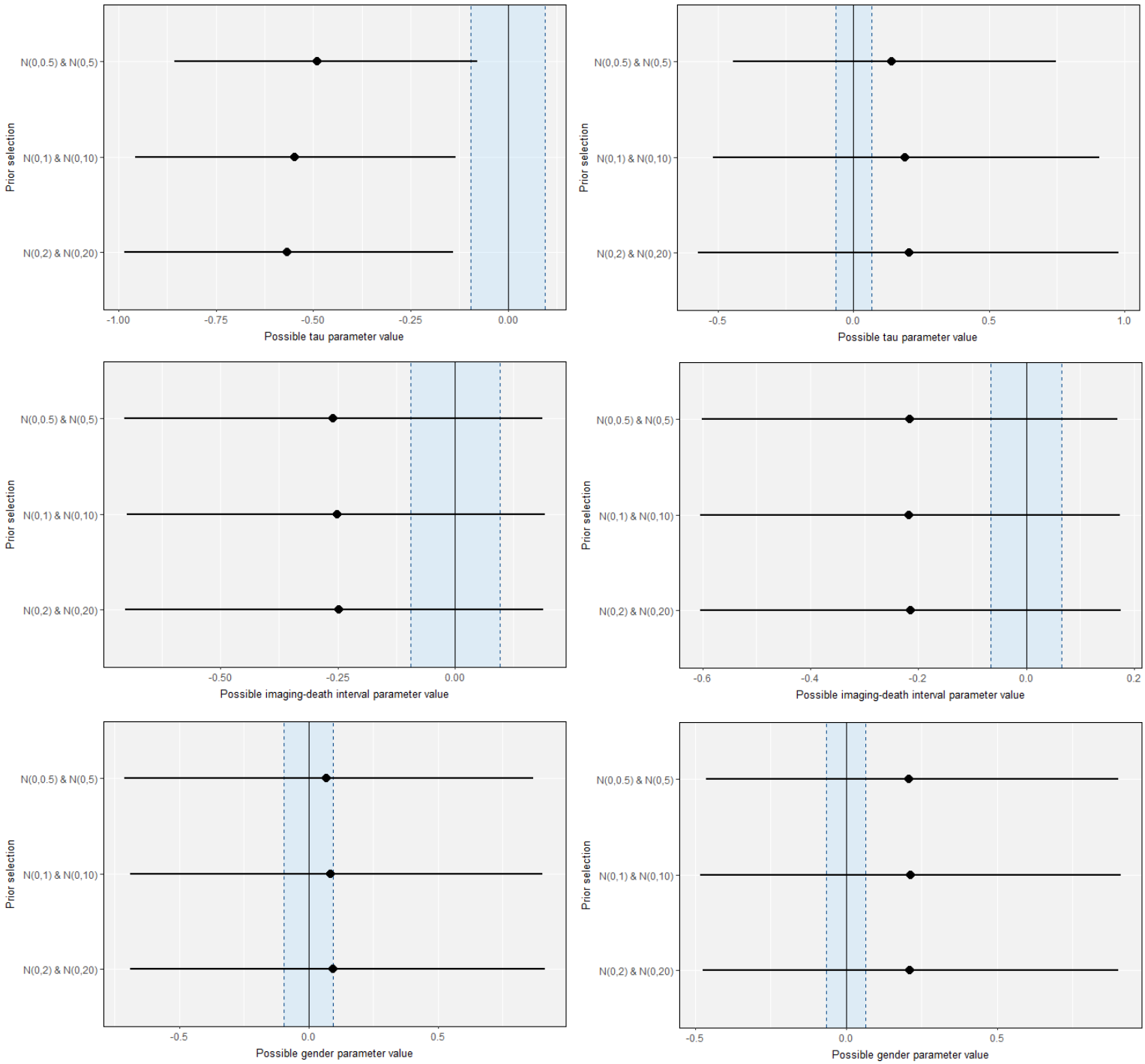



Figure 6.11: Sensitivity analysis plots of the effect of setting alternative weakly informative priors on the regression coefficient of the effect of interest in the CB density model in predicting atrophy of dentate nucleus with full dataset (left) and with a potential outlier removed (right). Normal distribution, N (mean, standard deviation), was chosen with mean centered at zero and standard deviation was varied. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE (blue region).

### 6.3.5. Less strongly interconnected regions accrue more tau burden

The mean weighted degree and closeness centrality of PSP subjects were assessed across matched cortical and subcortical regions (Figure 6.12). Occipital, cingulate, and primary motor regions showed higher weighted degree than other regions, with the globus pallidus showing the lowest weighted degree. In contrast, mean closeness centrality values between regions were similar, but was higher in the globus pallidus and pre-frontal regions as compared to other regions and was lowest in the occipital region.

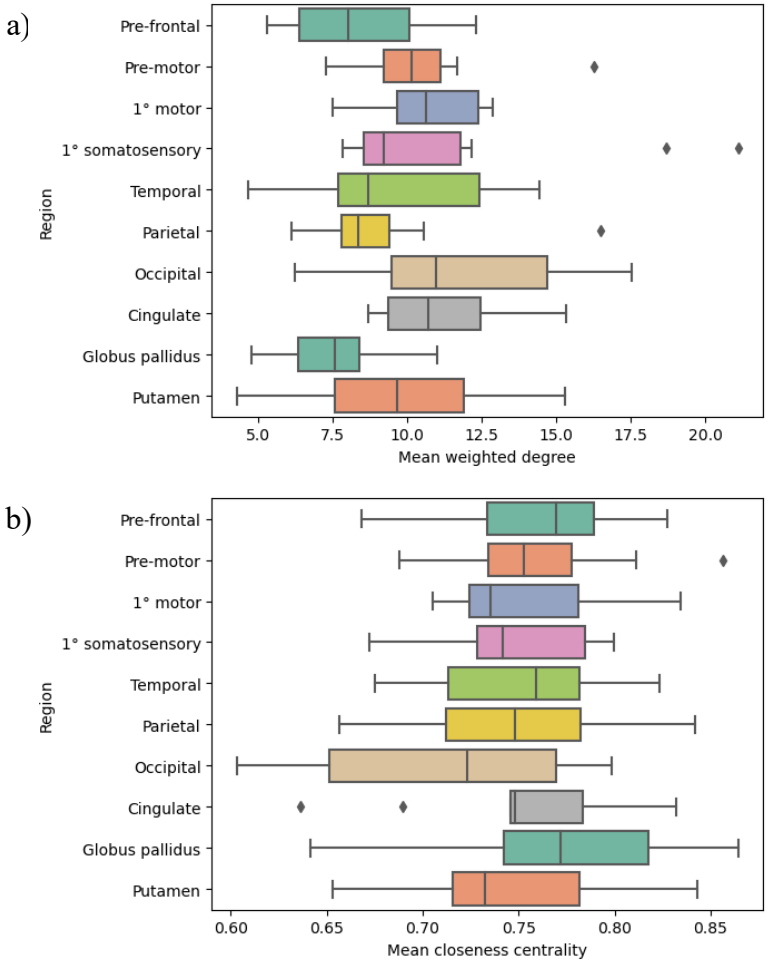



Figure 6.12: Boxplots showing a) mean weighted degree, b) closeness centrality in PSP subjects across all regions in the analysis.

Table 6.4: Bayes' factor (BF) for each tau type-specific model compared against the total tau density model in predicting weighted degree and closeness centrality across all regions and only cortical regions. *Indicates BF <1/3 or BF>3, illustrating moderate evidence in favor of total tau model or tau type-specific model, respectively. **indicates BF <1/10, illustrating strong evidence in favor of total tau model. 1/3< BF <3 represents inconclusive evidence for the total tau model as compared to tau type-specific model.

| Model | Weighted degree | Closeness centrality | Closeness centrality |
|---|---|---|---|
| Regions | All regions | All regions | Only cortical regions |
| Tau hallmark density | 0.87 | 2.12 | 3.19* |
| NFT density | 0.033** | 0.18* | 0.36 |
| TA density | 0.01** | 0.14* | 1.14 |
| CB density | 0.43 | 2.09 | 2.39 |
| TF density | 0.87 | 0.98 | 0.91 |

I assessed the weighted degree – tau relationship across *all brain regions* by constructing models with different tau density types and imaging-death interval, gender, echo acquisition and disease duration as covariates. At a 5% network threshold (Figure 6.13), there was a negative association between total tau density and weighted degree across all regions (median = -0.08, Crl -0.11 to -0.03, 0% in ROPE). To investigate whether tau type-specific burden is more informative of weighted degree than total tau burden, we assessed tau type-specific models against a total tau model (Table 6.4). There was strong evidence in favor of the total tau model against NFT and TA models, except for the TF, CB and tau hallmark models which showed inconclusive evidence against the total tau model. Results from Bayesian regression from 1% to 10% also consistently showed a negative association between total tau burden and weighted degree. Sensitivity analysis of the total tau model in predicting weighted degree (Figure 6.14) showed robust results where median of the posterior distribution of the regression coefficients stayed roughly unchanged.

Next, I assessed the closeness centrality – tau relationship across *all brain regions* in the same manner. At a 5% network threshold (Figure 6.13), there was insufficient evidence to support a positive association found between total tau density and closeness centrality across

all regions (median = 0.01, Crl 0.00 to 0.02, 9.58% in ROPE). To investigate whether tau type-specific burden is more informative of closeness centrality than total tau burden, we assessed tau type-specific models against a total tau model (Table 6.4). There was moderate evidence in favor of the total tau model against NFT and TA density, except for the TF, CB and tau hallmark models which showed inconclusive evidence against the total tau model. Results from Bayesian regression from 1% to 10% also consistently showed insufficient evidence to support the positive association between total tau burden and closeness centrality. Sensitivity analysis of the total tau model in predicting closeness centrality (Figure 6.15) showed robust results where median of the posterior distribution of the regression coefficients stayed roughly unchanged.

I further assessed closeness centrality – tau relationship across *only cortical regions*. At 5% network threshold, there was a positive trend between closeness centrality and tau but there was insufficient evidence supporting this association (median = 0.02, Crl 0.00 to 0.023 6.61% in ROPE). In contrast to tau type-specific models, tau hallmark burden was informative of closeness centrality than total tau burden (BF=3.19). Despite a positive association between tau hallmark density and closeness centrality, the evidence supporting this relationship was almost sufficient (median = 0.02, Crl 0.00 to 0.03 0.29% in ROPE). Results from 1-10% network thresholds showed consistently insufficient evidence for this relationship (0.30-27.65% in ROPE from 1-4% threshold, 0.23-2.13% in ROPE from 7-10% threshold), except for network threshold of 6% which showed sufficient evidence (median = 0.02, Crl 0.00 to 0.03, 0% in ROPE). Sensitivity analysis showed robust results where median of the posterior distribution of the regression coefficients stayed roughly unchanged.

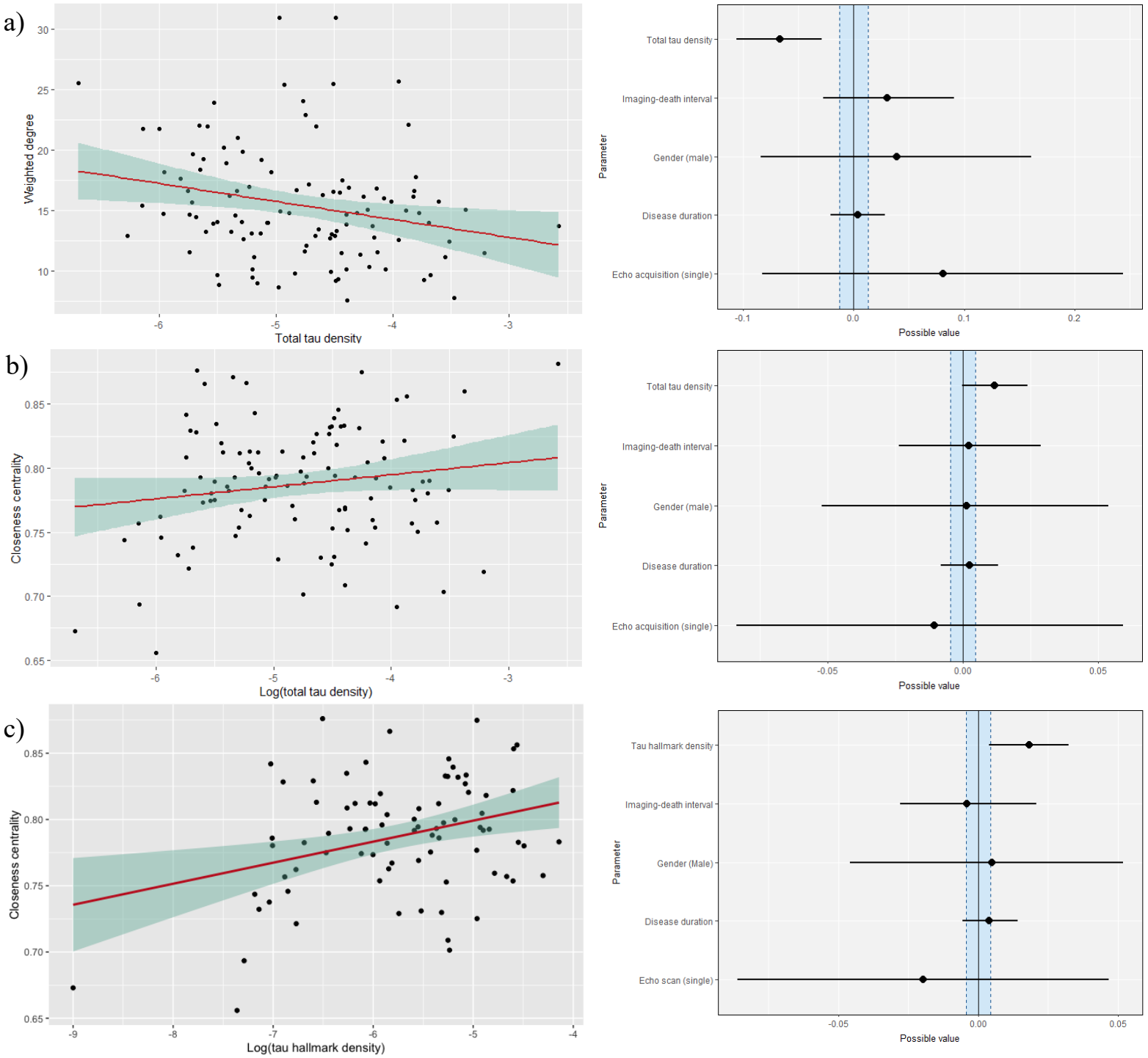



Figure 6.13: Scatterplots (left) showing total tau density and a) weighted degree or b) closeness centrality across all regions. c) Tau hallmark density and closeness centrality across only cortical regions is plotted. Plots (right) showing the posterior distribution of the regression coefficients of the tau models; graph metric ~ tau density + imaging-death interval + gender + disease duration + echo type + (1|region name) + (1|Patient ID). Mean (circle) and 95% credible interval are plotted with ROPE [-0.01 to 0.01] (blue region).

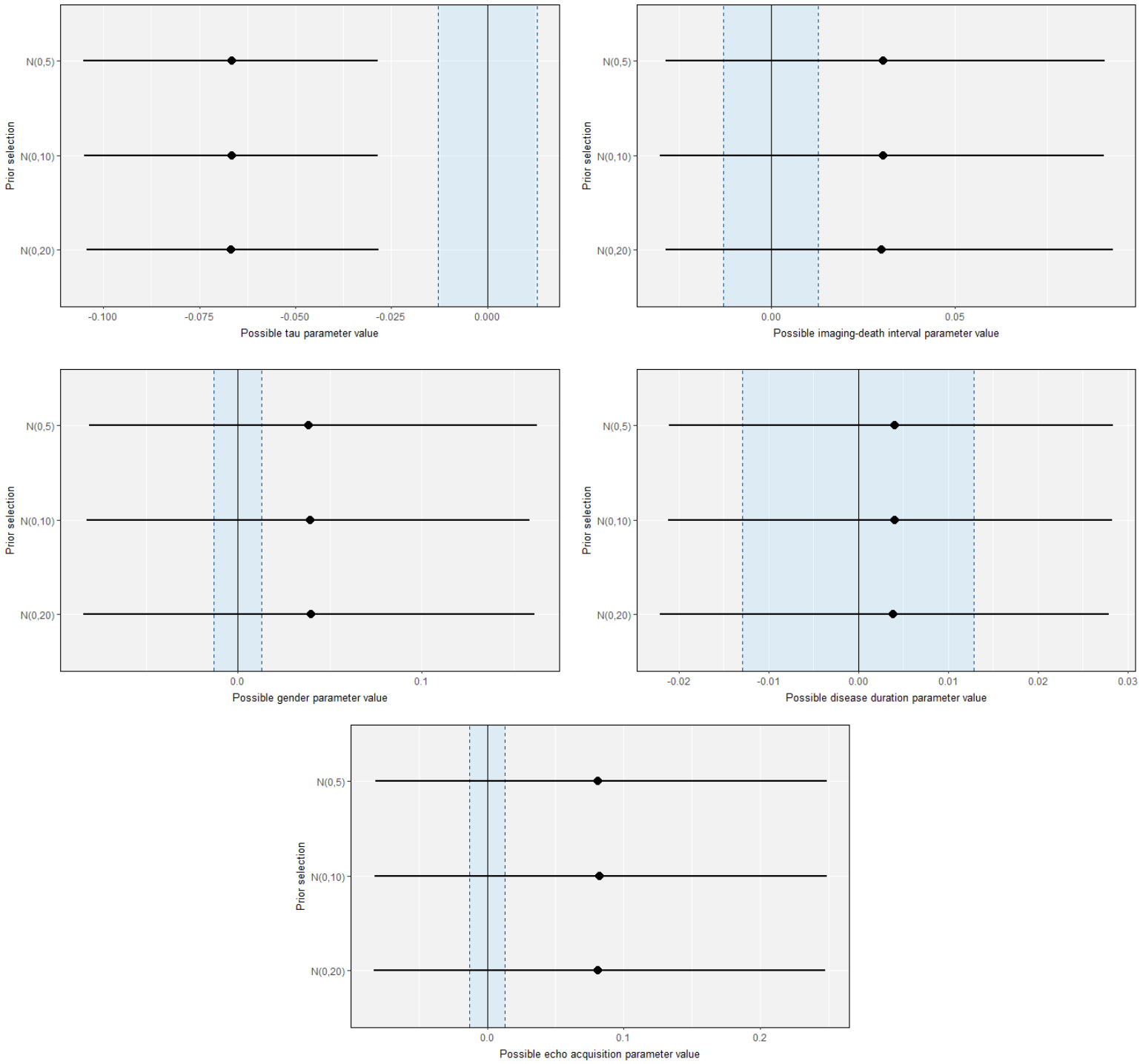



Figure 6.14: Sensitivity analysis plots of the effect of setting alternative weakly informative priors on the regression coefficient of the effect of interest in the total tau density model in predicting weighted degree across all regions. Normal distribution, N (mean, standard deviation), was chosen with mean centered at zero and standard deviation was varied. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE (blue region).

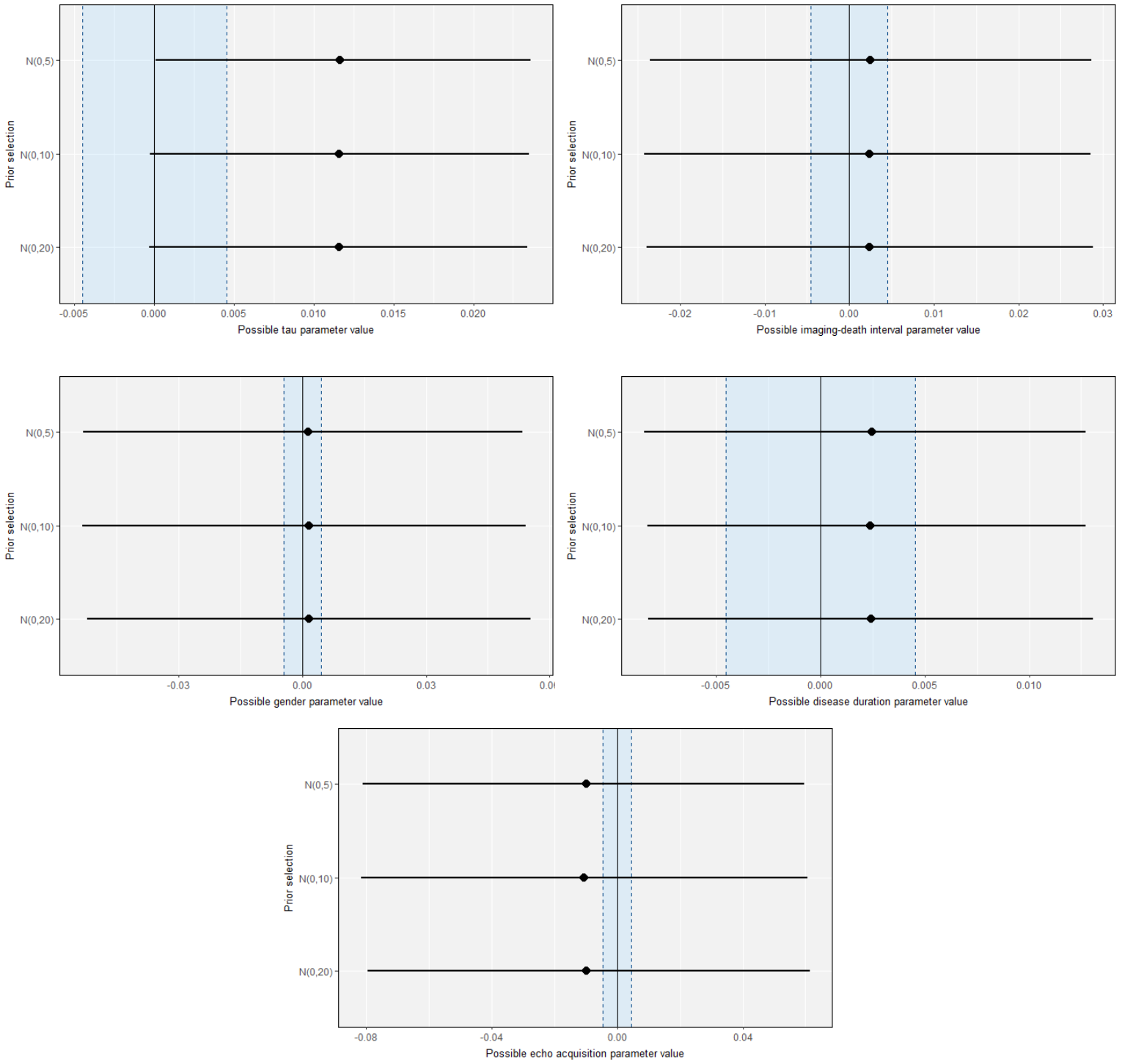



Figure 6.15: Sensitivity analysis plots of the effect of setting alternative weakly informative priors on the regression coefficient of the effect of interest in the total tau density model in predicting closeness centrality across all regions. Normal distribution, N (mean, standard deviation), was chosen with mean centered at zero, and standard deviation was varied. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE (blue region).

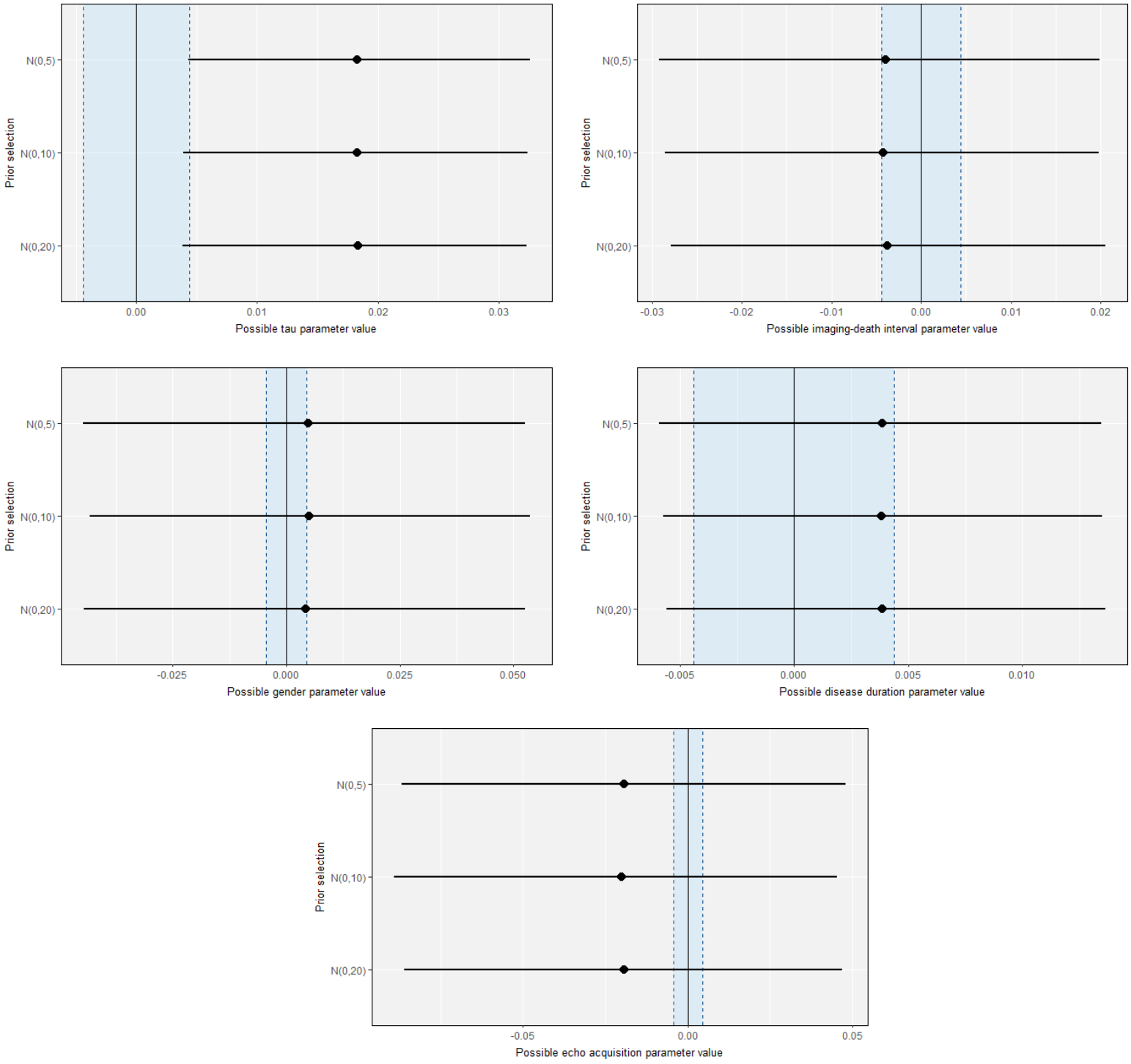



Figure 6.16: Sensitivity analysis plots of the effect of setting alternative weakly informative priors on the regression coefficient of the effect of interest in the tau hallmark density model in predicting closeness centrality across only cortical regions. Normal distribution, N (mean, standard deviation), was chosen with mean centered at zero and standard deviation was varied. Median (circle) and 95% credible interval (line) are plotted for each parameter alongside ROPE (blue region).

## 6.4. Discussion

### 6.4.1. Atrophy and tau burden

Our data suggests that PSP subjects showed the expected pattern of atrophy in subcortical regions such as the basal ganglia, thalamus (Albrecht et al., 2019) and brainstem nuclei and a wide range of cortical regions (Pan et al., 2017; Piattella et al., 2015; Stezin et al., 2017). Regions with marked atrophy included the midbrain and cerebellar pedunculi which are also typical of PSP (Albrecht et al., 2019; Höglinger et al., 2017; Pan et al., 2017).

In relation to atrophy-tau relationship, I found insufficient evidence to support that tau density is associated with structural atrophy at the whole-brain or cortical level. Nevertheless, there was a trend towards a negative relationship at the whole-brain level and a positive relationship when considering only cortical regions. The contrasting trends could potentially be explained by the differing severity of pathology in cortical and subcortical structures as tau primarily accrues in the basal ganglia before spreading to the cortex (Kovacs et al., 2020). This might suggest that pathology stage could differ between cortical and subcortical regions, highlighting the importance of assessing individual regions when considering the relationship between atrophy and neuropathological changes. Future studies will need to further validate these trends and hypotheses as I have found insufficient evidence to reject nor accept these findings.

In contrast to my findings, (Spina et al., 2019) found a positive relationship between tau burden and atrophy at the whole-brain level in PSP participants. Their effect size was very small and just reached significance ($b=0.036$, $p = 0.046$). With a few methodological differences in this study, it may not be surprising that the findings did not replicate. Spina et al (2019) included only PSP-RS subjects, regions from the right hemisphere and more subcortical regions such as the midbrain, amygdala, hippocampus, and thalamus in their study. Furthermore, another study (Carlos et al., 2022) investigated region-specific relationship between atrophy and tau burden in 4R-tauopathies (PSP and CBD combined). Consistent with our findings, they found no such relationship in the cortex. However, they found a positive atrophy-tau relationship in the subthalamic nucleus, but the influence of

CBD participants cannot be ruled out. Overall, our results suggest that atrophy derived measure from MRI does not reliably track tau burden in PSP.

As no atrophy-tau relationship was found in cortical regions where TA density was high, this could suggest a protective role of astrocytes (Carlos et al., 2022; Qian et al., 2023). It has been shown that astrocytes are unable to propagate tau in the absence of neuronal tau (Narasimhan et al., 2019), they may therefore accumulate tau secondary to neurons which may explain the lack of atrophy-tau relationship in cortical regions (Carlos et al., 2022). Bayesian model comparison showed that total tau density was equally as good as tau type-specific density as a predictor of atrophy across whole-brain and cortical regions. This is likely to be because tau does not correlate with atrophy, therefore the type of tau does not change the results.

I further investigated the atrophy-tau relationship within each brain region in the study and found a negative relationship between atrophy and CB density in the dentate nucleus. This relationship could become inconclusive if a narrower prior was chosen, and it could have been driven by an outlier as the relationship became inconclusive when it was removed. With other regions, there was insufficient evidence to support an atrophy-tau relationship, but positive and negative trends could be observed across the regions. In contrast to my findings, Carlos et al (2022) found a positive atrophy-tau relationship in the subthalamic nucleus but found no such relationship in cerebellar dentate nucleus. It is important to note a few methodological differences such that they investigated atrophy-tau relationship using brain volume and semi-quantitative measure of tau burden whereas I used w-score to indicate atrophy by taking age and gender into account, and digitally quantified tau density. Overall, my results suggest that atrophy is not a reliable marker of tau burden, even when assessing such relationship separately within each region.

### 6.4.2. Weighted degree & tau burden

I have demonstrated that regions with higher tau burden have lower weighted degree, or in other words are less functionally interconnected with other regions at the whole-brain level. This result is in the opposite direction to a previous study (Cope et al., 2018). There are two

major methodological differences that may explain the differences in findings, first, they found a positive relationship between global tau burden and weighted degree across subjects, but no relationship was found when region-specific tau burden was used. Also, they used tau PET which captures *in vivo* tau burden but also suffer from off-target binding issue in PSP, unlike *post mortem* tau burden (Lee et al., 2023). Therefore, future studies could aim to validate our findings using tau burden quantified from *post mortem* samples of other PSP cohorts.

Our results have implications for understanding tau spread hypotheses. Our finding does not support trans-neuronal spread of tau hypothesis since it would predict highly connected nodes to accrue higher tau burden as they are more likely to receive tau from the 'seed' regions (Cope et al., 2018). Our results are in-line with alternative hypotheses of tau where a node may be selectively vulnerable to tau pathology due to the lack of trophic support, higher metabolic demand or differential gene expression (Rittman et al., 2016; Zhou et al., 2012). Future studies could aim to disentangle these hypotheses by including more brain regions and exploring other graph metrics such as weighted participation coefficient as a proxy for metabolic activity and clustering coefficient as a proxy for trophic support (Cope et al., 2018).

Bayesian model comparison showed that total tau density is a better predictor of weighted degree than NFT and TA density but is no better than CB and TF density. This reinforces the benefit of assessing tau type-specific aggregates when assessing tau relationship with imaging measures. CB and TF are more abundant than NFT and TA density across brain regions in the study, which may suggest that the effect of tau burden on weighted degree is primarily driven by CB and TF density. The role of oligodendrocytes remains unclear in the pathogenesis of tauopathies, but a study (Narasimhan et al., 2019) has shown that oligodendrocytes can propagate tau independently of neuronal tau, unlike astrocytes in a mouse model. Neuroimaging studies have also shown that oligodendrocyte dysfunction can alter functional connectivity (Ji et al., 2017; Kawamura et al., 2020). In addition to oligodendrocytes being the most abundant cell type in the brain, this could explain why CB are the most abundant tau type in PSP. TF is more difficult to explain as it includes tau threads and small fragments of other tau aggregates. Future studies could further aim to improve the pipeline to cleanly detect tau threads to enable better understanding of TF density findings.

### 6.4.3. Closeness centrality and tau burden

There was insufficient evidence to support the association between closeness centrality and tau burden even though a positive trend was observed at both whole-brain and cortical levels. This is likely because closeness centrality values across brain regions in my analysis were relatively similar, unlike weighted degree.

Even though a positive association between closeness centrality and tau burden across cortical regions was found at 6% network density, the effect size was very small and insufficient evidence for the relationship was found across other thresholds in the 1-10% range. Therefore, it remains unclear whether tau burden is associated with closeness centrality as the posterior distribution of tau burden was also not completely within ROPE to permit the rejection of the alternative hypothesis. This may imply that a larger sample size, both number of subjects and brain regions are required to elucidate this relationship which may be more subtle than weighted degree and tau burden.

Bayesian model comparison when assessing closeness centrality and tau burden across cortical regions showed sufficient evidence in favor of tau hallmark burden than total tau burden. This likely suggests that a combined measure of TA, NFT and CB are useful in assessing closeness centrality, without the addition of TF burden which may introduce more noise. However, TF measure alone is not uninformative as model comparison suggests that each tau type-specific aggregate alone is as useful as total tau burden in relating to closeness centrality.

### 6.4.4. Overall

Although I have not compared them directly, my findings suggest that the functional MRI to tau relationship is stronger than the atrophy-tau relationship in PSP. This could be because tau first impairs the healthy functioning of cells which may directly or indirectly result in cell death (Sexton et al., 2022). Tau has been associated with synaptic loss at *post mortem* (Bigio et al., 2001) and *in vivo* using PET (Holland et al., 2020). As atrophy is a less direct effect of tau aggregation, it is possible that function-tau relationship better reflects these early neuropathological changes.

My findings also suggests that tau impairs cell functioning across the brain, while it may be possible that the atrophy-tau relationship is more focal, related to differing levels of severity of pathology across the brain and cell type affected. Overall, these results suggest that functional connectivity is more closely related to tau burden in PSP than structural measure.

### 6.4.5. Strengths and weaknesses

The main strength of this study is that tau burden was quantified digitally. This enabled tau type-specific quantification of PSP-related tau aggregates and detailed quantification such as tau density for each tau type. Second, the use of Bayesian statistics enabled the null hypothesis to be rejected, accepted or undecided which is useful for a study with relatively small sample size. The current study only included pathologically confirmed PSP cases, which ensured that the diagnosis was accurate.

Nevertheless, the matching between imaging-pathology samples is one of the main challenges in such study. It is not possible to perfectly match each *post mortem* brain slide which is a small thin section of the brain to the corresponding imaging sample. However, we tried to reduce noise or errors by averaging the imaging measures over several parcels. Moreover, *post mortem* samples, though provided detailed information of tau aggregates, can also only be taken at the time of death. This requires the inclusion of imaging-death interval to study the relationship between imaging-pathology samples, which in the context of linear regression, could be an oversimplification of the changes (Baayen et al., 2008). All participants with available fMRI and MRI data were included to maximise sample size in this study. This means participants with relatively longer imaging-death interval were also included where they may show a different imaging-pathology relationship to those with shorter interval. Brain volume estimation of subthalamic nucleus and dentate nucleus were carried out using probability maps as they are very small on an MRI scan. This means only estimated volumes, but not actual volumes could be extracted and that these structures are more prone to mis-localisation. Results concerning these regions should therefore be interpreted with caution. Furthermore, due to limited data, we were not able to assess the imaging-pathology relationship separately for subcortical regions, which could be interesting as PSP is a primarily subcortical disorder.

# Chapter 7: Discussion

## 7.1. Summary of the presented work

In this thesis, I set out to investigate the relationship between *in vivo* imaging markers and underlying tau pathology in PSP. I first attempted to advance the semi-quantitative approach of assessing pathology in *post mortem* brain by developing an automated pipeline to quantify tau positive and negative neuronal and glial cells using a supervised machine learning algorithm. I designed a pipeline by considering 3 main challenges associated with cell classification of cell class imbalance, the heterogeneity of cell characteristics and the ratio of class imbalance across brain regions, and the ambiguous nature of classifying cells. I compared different approaches for feature selection and machine learning algorithm. Balanced random forest with feature recursive elimination approach was selected for the final pipeline as it has the best trade-off accuracy, interpretability, and adaptability.

Despite these considerations, the cell quantification pipeline was unable to perform at a high level due to 3 main reasons. First, neuronal, and glial cells have overlapping features based on the information extracted from the brightfield image with haematoxylin staining. This issue could be because of the chosen feature extraction method or simply that our images only provided information on cell nuclei but no other important cytological features necessary to differentiate between the cell types. Secondly, the pipeline was not robust against tau aggregation as the nuclei detection step was inaccurate since nuclear shape was distorted by tau aggregates, resulting in misclassification of cell types. Thirdly, relying on DAB thresholding to identify tau positive cells is likely problematic due to the bleeding between digital channels.

I therefore concentrated on quantifying PSP-related tau pathological hallmarks which reflect PSP pathology. I adapted the original pipeline to be compatible with classifying 4 pathological hallmarks in key brain regions in the PSP pathology staging scheme: coiled bodies, neurofibrillary tangles, tufted astrocytes, and tau fragments. The pipeline has been shown to work at a much higher performance than the cell classification pipeline, comparable to an expert neuropathologist. Next, I validated the tau quantification pipeline in a larger

dataset consisting of PSP participants of pathological stage 2-6 and control donors. I found a strong correspondence between digitally quantified tau burden and the current consensus PSP staging scheme. I further assessed the relationship between tau burden, PSP stage and PSPRS score prior to death as a measure of clinical severity. I found that the clinical severity of PSP stage 6 participants is higher than that of stage 2, while clinical severity of stage 3-5 is similar to that of stage 2. Total cortical tau density and subcortical neurofibrillary tangle density, but not subcortical or whole-brain total tau density, correlate with the clinical severity. These results highlight the importance of studying tau type-specific burden in specific anatomical locations in PSP, instead of simply investigating total tau burden in all regions or using the pathological PSP stage. Next, I compared total tau density quantified from control to PSP participants and have shown that that former indeed had lower tau burden than the latter. This further suggests that the tau quantification pipeline also works well when tau is minimal and is not highly susceptible to noise from, for example, artefacts or bleeding between digital channels.

Lastly, I investigated the relationship between tau burden and both structural and functional imaging measures from PSP participants. I did not find evidence supporting the relationship between atrophy and tau burden across the whole brain or cortical regions. Nevertheless, when inspecting atrophy-tau relationship within each brain region, I found atrophy and coiled body density to be correlated only in dentate nucleus. Moreover, there was a negative relationship between functional connectivity and total tau burden across all brain regions in the study. These results therefore demonstrate that structural imaging measures may be a less reliable marker than functional imaging measures in tracking tau burden in PSP participants. Overall, I have advanced the semi-quantitative approach of assessing tau pathology in *post mortem* PSP brains and investigated the association between tau burden and a variety of measures such as PSP pathological staging, clinical severity, and imaging markers.

## 7.2. Relation to existing literature & future work

### 7.2.1. Digital pathology pipeline

In tauopathies, studies using automated pipeline have attempted to quantify neurons (Kaalund et al., 2020) and tau type-specific aggregates (Koga et al., 2021, 2022; Marx et al., 2022; Signaevsky et al., 2019). However, to my knowledge, no study has attempted to quantify both neuronal and glial cells as well as a wide range of tau type-specific aggregates in the same sample. Previous studies using automated pipelines can largely be grouped into 2 types, for mechanistic research (Kaalund et al., 2020; Marx et al., 2022; Signaevsky et al., 2019) and for aiding *post mortem* diagnosis of tauopathies (Koga et al., 2021, 2022). For mechanistic research, the relationship between overall neuronal and tau positive neuronal density to cognition has been investigated (Kaalund et al., 2020). Studies have also attempted to investigate diseases-specific tau distribution patterns in both grey and white matter but have not classified tau into different types (Coughlin et al., 2022; Vega et al., 2021). For studies that have assessed tau type-specific aggregates, the focus has been on investigating the relationship between neurofibrillary tangle and cognition (Marx et al., 2022; Signaevsky et al., 2019), and using disease-specific tau aggregates for *post mortem* diagnosis (Koga et al., 2021, 2022). Therefore, this thesis is the first, to my knowledge, to have successfully developed tau type-specific quantification pipeline for multiple key brain regions and investigated various association between tau burden and other factors such as the current pathological staging scheme, to cognition and imaging measures in one cohort.

Despite high performance of the tau quantification pipeline, it only recognizes PSP-related tau aggregates and would misclassify non-PSP-related aggregates in the presence of co-pathology as it has not 'seen' them before. Therefore, an obvious next step is to re-train the pipeline to recognize other tau pathological hallmarks (Jecmenica Lukic et al., 2020) such as Alzheimer's disease, and other tauopathies including Corticobasal Degeneration and Pick's disease. Eventually, it may be possible to build a pipeline that can distinguish multiple pathologies and recognize co-pathology. If successful, this pipeline could be applied to regions in PSP with heavy tau pathology but are known to have heavy co-pathology such as the midbrain (Kovacs et al., 2020). The pipeline could be trained to recognize other disease-specific non-tau protein hallmarks. This will require a few adaptations to the current object detection step as it is currently designed to detect tau objects from DAB thresholding.

However, the principles of the pipeline could be adapted for other pathological protein aggregates, such as alpha-synuclein, beta-amyloid or TDP-43.

### 7.2.2. Cell classification pipeline

Cell classification pipelines have received great attention in other areas of biological science research (Shifat-E-Rabbi et al., 2020), but less so in tauopathies as the focus has largely been on tau aggregates. Still, being able to quantify neuronal density as a proxy for neuronal loss, and glial density will continue to be highly desirable for investigating the relationship between tau aggregation and neuronal and glial cells. Studies investigating the interaction between tau and cells are mostly based on animal models of tauopathies where questions of tau hyperphosphorylation, aggregation and spreading have been addressed (Dujardin et al., 2015). An advantage of animal models is that they can provide clear experimental evidence for understanding how complex processes work where various hypotheses can be directly tested. This has led to findings including evidence supporting the toxicity of tau aggregates, leading to neuronal death (Fatouros et al., 2012; Mocanu et al., 2008), and evidence showing that cognitive impairment and neurodegeneration can happen without tangle formation (Andorfer et al., 2005; Cowan & Mudher, 2013; Wittmann et al., 2001). The limitation of animal models is that it is not clear whether these mechanisms contribute to human sporadic disease. As these studies are based on animal models which have translational potential to tauopathies in human, it would be interesting to investigate the association between cell density and tau aggregation in *post mortem* human samples.

Stereology has been the standard approach for cell counting but is time-intensive and require training (West & Gundersen, 1990). The isotropic fractionator approach was developed in 2005 and is a more user-friendly approach, but the tissue analyzed is destroyed and it provides no cellular spatial information (Herculano-Houzel & Lent, 2005). An automated approach promises to be an optimal alternative approach that provides rich information and is scalable. To improve the cell classification pipeline, we may require an alternative staining agent, such as a Nissl stain to identify cytoplasm and other cytological features, to enable more distinctive features of each class to be visible. Even if this were successful, the cell classification pipeline is still likely to require an additional tau object classification pipeline since cells with tau aggregates become distorted.

### 7.2.3. Pathological stage

I assessed the agreement between digitally quantified tau burden and the current PSP staging scheme which showed strong correspondence. This validates our novel digital approach and provide detailed insights into how the staging system works such as the key pathology in each region and the contribution of region-specific severity towards the overall PSP stage. The digital approach has additional advantages over the traditional semi-quantitative approach, providing more fine-grained detail over a much larger brain area and is less prone to inter and intra-rater variability.

I further assessed the predictive power of PSP stage, digital tau burden and PSPRS score using Bayesian regression. I found that cortical tau burden and subcortical neurofibrillary tangle burden are associated with PSPRS score, but not when tau was quantified from all regions. Previous studies have found a positive relationship between whole brain atrophy and PSPRS score (Dutt et al., 2016; Tsai et al., 2016), however, amongst all structures, the association is most prominent in the brainstem, midbrain and cortical regions such as the precentral gyrus (Dutt et al., 2016). Subcortical atrophy has also been shown to drive changes in frontal executive function (Whiteside et al., 2021), all together, these previous findings suggest that subcortical and cortical neuropathology burden measured in different ways underlie clinical dysfunction. In general, clinical deficits in neurodegenerative diseases are linked to neuronal dysfunction with unclear role of glial tau burden (Kovacs et al., 2020; Robinson et al., 2020), while this may appear to be the case in subcortical structures in our PSP cohort, such relationship is likely different in cortical structures where the interplay between neuronal and glial tau burden may be more prominent.

There was only sufficient evidence to support that clinical severity of PSP stage 6 is higher than stage 2 and not other stages. These results suggest using pathological stage as a measure of tau burden can mask subtle pathological changes where more detailed measurement such as tau density would be more informative, especially when linking to cognition (Signaevsky et al., 2019). This is in-line with a study which found Braak stage to have less clinicopathologic power than using neurofibrillary counts in PART (Marx et al., 2022). Directly using pathological counts would also avoid the need to validate such disease-specific staging scheme in other diseases with similar pathology which may have different pathogenesis.

### 7.2.4. In vivo imaging and its relation to tau burden

I investigated whether *in vivo* structural and functional imaging measures are indicative of tau burden using *post mortem* data. Such investigation is crucial for validating imaging measures as a surrogate biomarker of tau burden. I found that tau burden is associated with weighted degree but not atrophy across all brain regions in the PSP staging scheme. Despite the lack of atrophy-tau relationship in our PSP cohort, Spina et al (2019) found a small effect size of tau-atrophy relationship (that just reached significance) at the whole-brain level while Carlos et al (2021) found such relationship only in subcortical regions. Altogether, this suggests that it is possible to find no such relationship in other PSP cohorts with a few methodological differences. This also suggests that structural atrophy may not be a reliable biomarker of the underlying tau pathology in PSP. To my knowledge, this thesis is the first to investigate the relationship between tau burden and graph metrics in PSP cohort. Therefore, our finding remains to be replicated in other PSP cohorts to ensure that functional connectivity is a reliable biomarker of PSP tau pathology.

The investigation of imaging-pathology relationship also provides insights for understanding the relationship between tau burden, functional connectivity, and atrophy in tauopathies. My findings suggest that the relationship between tau and functional change is more closely related than structural change, in-line with the finding from chapter 5 (Page 124) that tau burden correlates with cognition. This is also consistent with the well-established insights that hyperphosphorylated tau can impair microtubule stability and affect normal functioning of neurons (Biernat et al., 2002; Irwin, 2016; Sexton et al., 2022; Y. Zhang et al., 2022). However, the relationship between tau burden and atrophy is less straightforward. In AD research, it is a common finding that structural MRI can be used to track NFT burden and can distinguish between AD subtypes (Apostolova et al., 2015; Burton et al., 2009; Jack et al., 2002; Whitwell et al., 2012). This therefore raises the question about the underlying basis of atrophy in neurodegenerative diseases, which is often interpreted as a marker of neuronal loss and neurodegeneration more generally (Jack et al., 2016). There are a few differences between AD and PS; beta-amyloid is also involved in AD, and the key tau pathology differs between the disorders. Future studies could try to investigate structural and functional imaging measures to *post mortem* tau burden in AD.

The relationship between tau burden and functional connectivity raises the possibility that the effect of tau neuropathology could be related to its role in supporting microtubules, and therefore the structural integrity of axons and synapses. This could be investigated by examining white matter. White matter in PSP also has substantial tau pathology but has received little attention. Studies have found disease-specific tau distribution patterns in grey and white matter (Coughlin et al., 2022; Vega et al., 2021). The current pipeline can easily be adapted to analyze tau pathology in white matter. This would permit comparison between white and grey matter in terms of tau burden, and relationship to clinical disease severity.

Autopsy-verified cohorts are essential in tauopathies as clinical diagnosis is challenging due to the overlapping clinical features between disorders which makes definitive diagnosis only possible at *post mortem*. For quantifying tau, *post mortem* data provides the most detailed quantification such as total and tau type-specific density. Though tau burden can be quantified using PET, it is only accurate in AD and not non-AD disorders due to off-target binding issues (Jucker & Walker, 2013). This means that *post mortem* tau is the current best method to quantify tau type-specific aggregates accurately. Concerns remain about matching between imaging and *post mortem* brain regions and making assumptions about how tau change with imaging-death interval need to be better addressed in future studies.

## 7.3. Concluding remarks

This PhD thesis has demonstrated the utility of machine learning in assessing neuropathology, specifically that an automated pipeline can achieve a fast and scalable workflow for quantifying tau pathology in PSP. Such a pipeline can be used to assess pathology across cell types and brain regions and is useful for mechanistic research in preliminarily assessing pathology and generating new hypotheses. The involvement of a well-trained neuropathologist will continue to be essential, but such a pipeline can assist in providing additional information and reducing workload. My work has highlighted the importance of studying tau type-specific burden in specific anatomical locations in PSP and the significance of investigating imaging-pathology relationships in both validating in *vivo* biomarkers and discovering novel mechanistic insights in the pathogenesis of tauopathies.

# Chapter 8: References

2018 Alzheimer's disease facts and figures. (2018). *Alzheimer's & Dementia*, *14*(3), 367–429. https://doi.org/10.1016/j.jalz.2018.02.001

Alafuzoff, I., Pikkarainen, M., Al-Sarraj, S., Arzberger, T., Bell, J., Bodi, I., Bogdanovic, N., Budka, H., Bugiani, O., Ferrer, I., Gelpi, E., Giaccone, G., Graeber, M. B., Hauw, J.-J., Kamphorst, W., King, A., Kopp, N., Korkolopoulou, P., Kovács, G. G., … Kretzschmar, H. (2006). Interlaboratory Comparison of Assessments of Alzheimer Disease-Related Lesions: A Study of the BrainNet Europe Consortium. *Journal of Neuropathology & Experimental Neurology*, *65*(8), 740–757. https://doi.org/10.1097/01.jnen.0000229986.17548.27

Alafuzoff, I., Thal, D. R., Arzberger, T., Bogdanovic, N., Al-Sarraj, S., Bodi, I., Boluda, S., Bugiani, O., Duyckaerts, C., Gelpi, E., Gentleman, S., Giaccone, G., Graeber, M., Hortobagyi, T., Höftberger, R., Ince, P., Ironside, J. W., Kavantzas, N., King, A., … Kretzschmar, H. (2009). Assessment of β-amyloid deposits in human brain: A study of the BrainNet Europe Consortium. *Acta Neuropathologica*, *117*(3), 309–320. https://doi.org/10.1007/s00401-009-0485-4

Albrecht, F., Bisenius, S., Neumann, J., Whitwell, J., & Schroeter, M. L. (2019). Atrophy in midbrain & cerebral/cerebellar pedunculi is characteristic for progressive supranuclear palsy – A double-validation whole-brain meta-analysis. *NeuroImage: Clinical*, *22*, 101722. https://doi.org/10.1016/j.nicl.2019.101722

Al-Janabi, S., Huisman, A., & Van Diest, P. J. (2012). Digital pathology: Current status and future perspectives. *Histopathology*, *61*(1), 1–9. https://doi.org/10.1111/j.1365-2559.2011.03814.x

Andorfer, C., Acker, C. M., Kress, Y., Hof, P. R., Duff, K., & Davies, P. (2005). Cell-cycle reentry and cell death in transgenic mice expressing nonmutant human tau isoforms. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(22), 5446–5454. https://doi.org/10.1523/JNEUROSCI.4637-04.2005

Andreadis, A. (2005). Tau gene alternative splicing: Expression patterns, regulation and modulation of function in normal brain and neurodegenerative diseases. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, *1739*(2), 91–103. https://doi.org/10.1016/j.bbadis.2004.08.010

Apostolova, L. G., Zarow, C., Biado, K., Hurtz, S., Boccardi, M., Somme, J., Honarpisheh, H., Blanken, A. E., Brook, J., Tung, S., Kraft, E., Lo, D., Ng, D., Alger, J. R., Vinters, H. V., Bocchetta, M., Duvernoy, H., Jack Jr., C. R., Frisoni, G. B., … Watson, C. (2015). Relationship between hippocampal atrophy and neuropathology markers: A 7T MRI validation study of the EADC-ADNI Harmonized Hippocampal Segmentation Protocol. *Alzheimer's & Dementia*, *11*(2), 139–150. https://doi.org/10.1016/j.jalz.2015.01.001

Armstrong, M. J., Castellani, R. J., & Reich, S. G. (2014). "Rapidly" Progressive Supranuclear Palsy. *Movement Disorders Clinical Practice*, *1*(1), 70–72. https://doi.org/10.1002/mdc3.12011

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017). QuPath: Open source software for digital

pathology image analysis. *Scientific Reports*, *7*(1), Article 1. https://doi.org/10.1038/s41598-017-17204-5

Basha, S. J., Madala, S. R., Vivek, K., Kumar, E. S., & Ammannamma, T. (2022). A Review on Imbalanced Data Classification Techniques. *2022 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, 1–6. https://doi.org/10.1109/ICACTA54488.2022.9753392

Bautista, P. A., Hashimoto, N., & Yagi, Y. (2014). Color standardization in whole slide imaging using a color calibration slide. *Journal of Pathology Informatics*, *5*(1), 4. https://doi.org/10.4103/2153-3539.126153

Bejanin, A., Schonhaut, D. R., La Joie, R., Kramer, J. H., Baker, S. L., Sosa, N., Ayakta, N., Cantwell, A., Janabi, M., Lauriola, M., O'Neil, J. P., Gorno-Tempini, M. L., Miller, Z. A., Rosen, H. J., Miller, B. L., Jagust, W. J., & Rabinovici, G. D. (2017). Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. *Brain*, *140*(12), 3286–3300. https://doi.org/10.1093/brain/awx243

Bennett, D. A., Schneider, J. A., Arvanitakis, Z., Kelly, J. F., Aggarwal, N. T., Shah, R. C., & Wilson, R. S. (2006). Neuropathology of older persons without cognitive impairment from two community-based studies. *Neurology*, *66*(12), 1837–1844. https://doi.org/10.1212/01.wnl.0000219668.47116.e6

Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V., & Madabhushi, A. (2019). Artificial intelligence in digital pathology—New tools for diagnosis and precision oncology. *Nature Reviews Clinical Oncology*, *16*(11), Article 11. https://doi.org/10.1038/s41571-019-0252-y

Biernat, J., Wu, Y.-Z., Timm, T., Zheng-Fischhöfer, Q., Mandelkow, E., Meijer, L., & Mandelkow, E.-M. (2002). Protein Kinase MARK/PAR-1 Is Required for Neurite

Outgrowth and Establishment of Neuronal Polarity. *Molecular Biology of the Cell*, *13*(11), 4013–4028. https://doi.org/10.1091/mbc.02-03-0046

Bigio, E. H., Vono, M. B., Satumtira, S., Adamson, J., Sontag, E., Hynan, L. S., White, C. L., Baker, M., & Hutton, M. (2001). Cortical synapse loss in progressive supranuclear palsy. *Journal of Neuropathology and Experimental Neurology*, *60*(5), 403–410. https://doi.org/10.1093/jnen/60.5.403

Boulesteix, A.-L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *WIREs Data Mining and Knowledge Discovery*, *2*(6), 493–507. https://doi.org/10.1002/widm.1072

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

Braak, H., & Braak, E. (1995). Staging of alzheimer's disease-related neurofibrillary changes. *Neurobiology of Aging*, *16*(3), 271–278. https://doi.org/10.1016/0197-4580(95)00021-6

Braak, H., Thal, D. R., Ghebremedhin, E., & Del Tredici, K. (2011). Stages of the pathologic process in Alzheimer disease: Age categories from 1 to 100 years. *Journal of Neuropathology and Experimental Neurology*, *70*(11), 960–969. https://doi.org/10.1097/NEN.0b013e318232a379

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Briggs, M., Allinson, K. S., Malpetti, M., Spillantini, M. G., Rowe, J. B., & Kaalund, S. S. (2021). Validation of the new pathology staging system for progressive supranuclear palsy. *medRxiv*, 2021.01.18.21250017. https://doi.org/10.1101/2021.01.18.21250017

Brownlee, J. (2020). *Imbalanced classification with Python: Better metrics, balance skewed classes, cost-sensitive learning*. Machine Learning Mastery.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), Article 3. https://doi.org/10.1038/nrn2575

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395–411.

Bürkner, P.-C. (2021). Bayesian Item Response Modeling in R with brms and Stan. *Journal of Statistical Software*, *100*, 1–54. https://doi.org/10.18637/jss.v100.i05

Burton, E. J., Barber, R., Mukaetova-Ladinska, E. B., Robson, J., Perry, R. H., Jaros, E., Kalaria, R. N., & O'Brien, J. T. (2009). Medial temporal lobe atrophy on MRI differentiates Alzheimer's disease from dementia with Lewy bodies and vascular cognitive impairment: A prospective study with pathological verification of diagnosis. *Brain*, *132*(1), 195–203. https://doi.org/10.1093/brain/awn298

Carlos, A. F., Tosakulwong, N., Weigand, S. D., Buciuc, M., Ali, F., Clark, H. M., Botha, H., Utianski, R. L., Machulda, M. M., Schwarz, C. G., Reid, R. I., Senjem, M. L., Jack, C. R., Ahlskog, J. E., Dickson, D. W., Josephs, K. A., & Whitwell, J. L. (2022). Histologic lesion type correlates of magnetic resonance imaging biomarkers in four-repeat tauopathies. *Brain Communications*, *4*(3), fcac108. https://doi.org/10.1093/braincomms/fcac108

Chen, C. (2011). *Using Random Forest to Learn Imbalanced Data*. 12.

Chételat, G., Landeau, B., Eustache, F., Mézenge, F., Viader, F., de la Sayette, V., Desgranges, B., & Baron, J.-C. (2005). Using voxel-based morphometry to map the

structural changes associated with rapid conversion in MCI: A longitudinal MRI study. *NeuroImage*, *27*(4), 934–946. https://doi.org/10.1016/j.neuroimage.2005.05.015

Cho, H., Choi, J. Y., Hwang, M. S., Kim, Y. J., Lee, H. M., Lee, H. S., Lee, J. H., Ryu, Y. H., Lee, M. S., & Lyoo, C. H. (2016). In vivo cortical spreading pattern of tau and amyloid in the Alzheimer disease spectrum. *Annals of Neurology*, *80*(2), 247–258. https://doi.org/10.1002/ana.24711

Chouliaras, L., & O'Brien, J. T. (2023). The use of neuroimaging techniques in the early and differential diagnosis of dementia. *Molecular Psychiatry*, 1–14. https://doi.org/10.1038/s41380-023-02215-8

Chung, D. C., Roemer, S., Petrucelli, L., & Dickson, D. W. (2021). Cellular and pathological heterogeneity of primary tauopathies. *Molecular Neurodegeneration*, *16*(1), 57. https://doi.org/10.1186/s13024-021-00476-x

Cohen, T. J., Guo, J. L., Hurtado, D. E., Kwong, L. K., Mills, I. P., Trojanowski, J. Q., & Lee, V. M. Y. (2011). The acetylation of tau inhibits its function and promotes pathological tau aggregation. *Nature Communications*, *2*(1), Article 1. https://doi.org/10.1038/ncomms1255

Colin, M., Dujardin, S., Schraen-Maschke, S., Meno-Tetang, G., Duyckaerts, C., Courade, J.-P., & Buée, L. (2020). From the prion-like propagation hypothesis to therapeutic strategies of anti-tau immunotherapy. *Acta Neuropathologica*, *139*(1), 3–25. https://doi.org/10.1007/s00401-019-02087-9

Conrad, C., & Gerlich, D. W. (2010). Automated microscopy for high-content RNAi screening. *The Journal of Cell Biology*, *188*(4), 453–461. https://doi.org/10.1083/jcb.200910105

Cope, T. E., Rittman, T., Borchert, R. J., Jones, P. S., Vatansever, D., Allinson, K., Passamonti, L., Vazquez Rodriguez, P., Bevan-Jones, W. R., O'Brien, J. T., & Rowe, J. B. (2018). Tau burden and the functional connectome in Alzheimer's disease and progressive supranuclear palsy. *Brain*, *141*(2), 550–567. https://doi.org/10.1093/brain/awx347

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Coughlin, D. G., Hiniker, A., Peterson, C., Kim, Y., Arezoumandan, S., Giannini, L., Pizzo, D., Weintraub, D., Siderowf, A., Litvan, I., Rissman, R. A., Galasko, D., Hansen, L., Trojanowski, J. Q., Lee, E., Grossman, M., & Irwin, D. (2022). Digital Histological Study of Neocortical Grey and White Matter Tau Burden Across Tauopathies. *Journal of Neuropathology & Experimental Neurology*, nlac094. https://doi.org/10.1093/jnen/nlac094

Cowan, C., & Mudher, A. (2013). Are Tau Aggregates Toxic or Protective in Tauopathies? *Frontiers in Neurology*, *4*. https://www.frontiersin.org/articles/10.3389/fneur.2013.00114

Cunningham, P., Cord, M., & Delany, S. J. (2008). Supervised Learning. In M. Cord & P. Cunningham (Eds.), *Machine Learning Techniques for Multimedia: Case Studies on Organization and Retrieval* (pp. 21–49). Springer. https://doi.org/10.1007/978-3-540-75171-7_2

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. *Proceedings of the 23rd International Conference on Machine Learning*, 233–240. https://doi.org/10.1145/1143844.1143874

Del Tredici, K., & Braak, H. (2020). To stage, or not to stage. *Current Opinion in Neurobiology*, *61*, 10–22. https://doi.org/10.1016/j.conb.2019.11.008

Dugger, B. N., & Dickson, D. W. (2017). Pathology of Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology*, *9*(7). https://doi.org/10.1101/cshperspect.a028035

Dujardin, S., Colin, M., & Buée, L. (2015). Invited review: Animal models of tauopathies and their implications for research/translation into the clinic. *Neuropathology and Applied Neurobiology*, *41*(1), 59–80. https://doi.org/10.1111/nan.12200

Dutt, S., Binney, R. J., Heuer, H. W., Luong, P., Attygalle, S., Bhatt, P., Marx, G. A., Elofson, J., Tartaglia, M. C., Litvan, I., McGinnis, S. M., Dickerson, B. C., Kornak, J., Waltzman, D., Voltarelli, L., Schuff, N., Rabinovici, G. D., Kramer, J. H., Jack, C. R., … Boxer, A. L. (2016). Progression of brain atrophy in PSP and CBS over 6 months and 1 year. *Neurology*, *87*(19), 2016–2025. https://doi.org/10.1212/WNL.0000000000003305

Elble, R. (2021, January 2). *Progressive Supranuclear Palsy—Symptoms, Causes, Treatment | NORD*. https://rarediseases.org/rare-diseases/progressive-supranuclear-palsy/

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A. R., Fox, P. T., Eickhoff, S. B., Yu, C., & Jiang, T. (2016). The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. *Cerebral Cortex*, *26*(8), 3508–3526. https://doi.org/10.1093/cercor/bhw157

Fatouros, C., Pir, G. J., Biernat, J., Koushika, S. P., Mandelkow, E., Mandelkow, E.-M., Schmidt, E., & Baumeister, R. (2012). Inhibition of tau aggregation in a novel Caenorhabditis elegans model of tauopathy mitigates proteotoxicity. *Human Molecular Genetics*, *21*(16), 3587–3603. https://doi.org/10.1093/hmg/dds190

Faye Carlos, A., Tosakulwong, N., Weigand, S. D., Buciuc, M., Ali, F., Clark, H. M., Botha, H., Utianski, R. L., Machulda, M. M., Schwarz, C. G., Reid, R. I., Senjem, M. L., Jack, C. R., Ahlskog, J. E., Dickson, D. W., Josephs, K. A., & Whitwell, J. L. (2022).

Histologic lesion type correlates of magnetic resonance imaging biomarkers in 4-repeat tauopathies. *Brain Communications*, fcac108. https://doi.org/10.1093/braincomms/fcac108

Fischl, B., & Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11050–11055.

Forrest, S. L., Kril, J. J., & Halliday, G. M. (2019). Cellular and regional vulnerability in frontotemporal tauopathies. *Acta Neuropathologica*, *138*(5), 705–727. https://doi.org/10.1007/s00401-019-02035-7

Franzmeier, N., Dewenter, A., Frontzkowski, L., Dichgans, M., Rubinski, A., Neitzel, J., Smith, R., Strandberg, O., Ossenkoppele, R., Buerger, K., Duering, M., Hansson, O., & Ewers, M. (2020). Patient-centered connectivity-based prediction of tau pathology spread in Alzheimer's disease. *Science Advances*, *6*(48), eabd1327. https://doi.org/10.1126/sciadv.abd1327

Frisoni, G. B., Fox, N. C., Jack, C. R., Scheltens, P., & Thompson, P. M. (2010). The clinical use of structural MRI in Alzheimer disease. *Nature Reviews Neurology*, *6*(2), Article 2. https://doi.org/10.1038/nrneurol.2009.215

Gao, Z., Wang, L., Zhou, L., & Zhang, J. (2017). HEp-2 Cell Image Classification With Deep Convolutional Neural Networks. *IEEE Journal of Biomedical and Health Informatics*, *21*(2), 416–428. https://doi.org/10.1109/JBHI.2016.2526603

Garcia, J. O., Ashourvan, A., Muldoon, S. F., Vettel, J. M., & Bassett, D. S. (2018). Applications of community detection techniques to brain graphs: Algorithmic considerations and implications for neural function. *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, *106*(5), 846–867. https://doi.org/10.1109/JPROC.2017.2786710

Giannakopoulos, P., Herrmann, F. R., Bussière, T., Bouras, C., Kövari, E., Perl, D. P., Morrison, J. H., Gold, G., & Hof, P. R. (2003). Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology*, *60*(9), 1495–1500. https://doi.org/10.1212/01.WNL.0000063311.58879.01

Grundman, M., Sencakova, D., Jack, C. R., Petersen, R. C., Kim, H. T., Schultz, A., Weiner, M. F., DeCarli, C., DeKosky, S. T., van Dyck, C., Thomas, R. G., Thal, L. J., & the Alzheimer's Disease Cooperative Study. (2002). Brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial. *Journal of Molecular Neuroscience*, *19*(1), 23–27. https://doi.org/10.1007/s12031-002-0006-6

Guo, J. L., & Lee, V. M. Y. (2014). Cell-to-cell transmission of pathogenic proteins in neurodegenerative diseases. *Nature Medicine*, *20*(2), Article 2. https://doi.org/10.1038/nm.3457

Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). *Feature Extraction: Foundations and Applications*. Springer.

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, *46*(1), 389–422. https://doi.org/10.1023/A:1012487302797

Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). *Exploring Network Structure, Dynamics, and Function using NetworkX*.

Hamilton, P. W., Bankhead, P., Wang, Y., Hutchinson, R., Kieran, D., McArt, D. G., James, J., & Salto-Tellez, M. (2014). Digital pathology and image analysis in tissue biomarker research. *Methods*, *70*(1), 59–73. https://doi.org/10.1016/j.ymeth.2014.06.015

Harper, L., Fumagalli, G. G., Barkhof, F., Scheltens, P., O'Brien, J. T., Bouwman, F., Burton, E. J., Rohrer, J. D., Fox, N. C., Ridgway, G. R., & Schott, J. M. (2016). MRI visual

rating scales in the diagnosis of dementia: Evaluation in 184 post-mortem confirmed cases. *Brain*, *139*(4), 1211–1225. https://doi.org/10.1093/brain/aww005

Herculano-Houzel, S., & Lent, R. (2005). Isotropic Fractionator: A Simple, Rapid Method for the Quantification of Total Cell and Neuron Numbers in the Brain. *The Journal of Neuroscience*, *25*(10), 2518–2521. https://doi.org/10.1523/JNEUROSCI.4526-04.2005

Hoenig, M. C., Bischof, G. N., Seemiller, J., Hammes, J., Kukolja, J., Onur, Ö. A., Jessen, F., Fliessbach, K., Neumaier, B., Fink, G. R., van Eimeren, T., & Drzezga, A. (2018). Networks of tau distribution in Alzheimer's disease. *Brain: A Journal of Neurology*, *141*(2), 568–581. https://doi.org/10.1093/brain/awx353

Höglinger, G. U., Respondek, G., & Kovacs, G. G. (2018). New classification of tauopathies. *Revue Neurologique*, *174*(9), 664–668. https://doi.org/10.1016/j.neurol.2018.07.001

Höglinger, G. U., Respondek, G., Stamelou, M., Kurz, C., Josephs, K. A., Lang, A. E., Mollenhauer, B., Müller, U., Nilsson, C., Whitwell, J. L., Arzberger, T., Englund, E., Gelpi, E., Giese, A., Irwin, D. J., Meissner, W. G., Pantelyat, A., Rajput, A., van Swieten, J. C., … for the Movement Disorder Society-endorsed PSP Study Group. (2017). Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria: MDS Clinical Diagnostic Criteria for PSP. *Movement Disorders*, *32*(6), 853–864. https://doi.org/10.1002/mds.26987

Holland, N., Jones, P. S., Savulich, G., Wiggins, J. K., Hong, Y. T., Fryer, T. D., Manavaki, R., Sephton, S. M., Boros, I., Malpetti, M., Hezemans, F. H., Aigbirhio, F. I., Coles, J. P., O'Brien, J., & Rowe, J. B. (2020). Synaptic loss in primary tauopathies revealed by [11C]UCB-J positron emission tomography. *Movement Disorders : Official Journal of the Movement Disorder Society*, *35*(10), 1834–1842. https://doi.org/10.1002/mds.28188

Hsu, C.-W., Chang, C.-C., Lin, C.-J., & others. (2003). *A practical guide to support vector classification*. Taipei, Taiwan.

Iba, M., Guo, J. L., McBride, J. D., Zhang, B., Trojanowski, J. Q., & Lee, V. M.-Y. (2013). Synthetic tau fibrils mediate transmission of neurofibrillary tangles in a transgenic mouse model of Alzheimer's-like tauopathy. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(3), 1024–1037. https://doi.org/10.1523/JNEUROSCI.2642-12.2013

Iida, M. A., Farrell, K., Walker, J. M., Richardson, T. E., Marx, G. A., Bryce, C. H., Purohit, D., Ayalon, G., Beach, T. G., Bigio, E. H., Cortes, E. P., Gearing, M., Haroutunian, V., McMillan, C. T., Lee, E. B., Dickson, D. W., McKee, A. C., Stein, T. D., Trojanowski, J. Q., … Crary, J. F. (2021). Predictors of cognitive impairment in primary age-related tauopathy: An autopsy study. *Acta Neuropathologica Communications*, *9*(1), 134. https://doi.org/10.1186/s40478-021-01233-3

Ingala, S., Tomassen, J., Collij, L. E., Prent, N., van 't Ent, D., ten Kate, M., Konijnenberg, E., Yaqub, M., Scheltens, P., de Geus, E. J. C., Teunissen, C. E., Tijms, B., Wink, A. M., Barkhof, F., van Berckel, B. N. M., Visser, P. J., & den Braber, A. (2021). Amyloid-driven disruption of default mode network connectivity in cognitively healthy individuals. *Brain Communications*, *3*(4), fcab201. https://doi.org/10.1093/braincomms/fcab201

Irwin, D. J. (2016). Tauopathies as clinicopathological entities. *Parkinsonism & Related Disorders*, *22*, S29–S33. https://doi.org/10.1016/j.parkreldis.2015.09.020

Iturria-Medina, Y., Sotero, R. C., Toussaint, P. J., Mateos-Pérez, J. M., & Evans, A. C. (2016). Early role of vascular dysregulation on late-onset Alzheimer's disease based on multifactorial data-driven analysis. *Nature Communications*, *7*(1), 11934. https://doi.org/10.1038/ncomms11934

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Dunn, B., Haeberlein, S. B., Holtzman, D. M., Jagust, W., Jessen, F., Karlawish, J., Liu, E., Molinuevo, J. L., Montine, T., Phelps, C., Rankin, K. P., Rowe, C. C., Scheltens, P., Siemers, E., Snyder, H. M., … Silverberg, N. (2018). NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, *14*(4), 535–562. https://doi.org/10.1016/j.jalz.2018.02.018

Jack, C. R., Bennett, D. A., Blennow, K., Carrillo, M. C., Feldman, H. H., Frisoni, G. B., Hampel, H., Jagust, W. J., Johnson, K. A., Knopman, D. S., Petersen, R. C., Scheltens, P., Sperling, R. A., & Dubois, B. (2016). A/T/N: An unbiased descriptive classification scheme for Alzheimer disease biomarkers. *Neurology*, *87*(5), 539–547. https://doi.org/10.1212/WNL.0000000000002923

Jack, C. R., Dickson, D. W., Parisi, J. E., Xu, Y. C., Cha, R. H., O'Brien, P. C., Edland, S. D., Smith, G. E., Boeve, B. F., Tangalos, E. G., Kokmen, E., & Petersen, R. C. (2002). Antemortem MRI findings correlate with hippocampal neuropathology in typical aging and dementia. *Neurology*, *58*(5), 750–757. https://doi.org/10.1212/WNL.58.5.750

Jack, C. R., Petersen, R. C., Xu, Y. C., Waring, S. C., O'Brien, P. C., Tangalos, E. G., Smith, G. E., Ivnik, R. J., & Kokmen, E. (1997). Medial Temporal Atrophy on MRI in Normal Aging and Very Mild Alzheimer's Disease. *Neurology*, *49*(3), 786–794.

Jardim, A. P., Corso, J. T., Garcia, M. T. F. C., Gaça, L. B., Comper, S. M., Lancellotti, C. L. P., Centeno, R. S., Carrete, H., Cavalheiro, E. A., Scorza, C. A., & Yacubian, E. M. T. (2016). Hippocampal atrophy on MRI is predictive of histopathological patterns and surgical prognosis in mesial temporal lobe epilepsy with hippocampal sclerosis. *Epilepsy Research*, *128*, 169–175. https://doi.org/10.1016/j.eplepsyres.2016.10.014

Jecmenica Lukic, M., Kurz, C., Respondek, G., Grau-Rivera, O., Compta, Y., Gelpi, E., Troakes, C., Barcelona Brain Bank collaborative group, the MDS-endorsed PSP study group, van Swieten, J. C., Giese, A., Roeber, S., Arzberger, T., & Höglinger, G. (2020). Copathology in Progressive Supranuclear Palsy: Does It Matter? *Movement Disorders: Official Journal of the Movement Disorder Society*, *35*(6), 984–993. https://doi.org/10.1002/mds.28011

Jellinger, K. A. (2010). Con: Can neuropathology really confirm the exact diagnosis? *Alzheimer's Research & Therapy*, *2*(2), 11. https://doi.org/10.1186/alzrt34

Jellinger, K. A. (2018). Different patterns of hippocampal tau pathology in Alzheimer's disease and PART. *Acta Neuropathologica*, *136*(5), 811–813. https://doi.org/10.1007/s00401-018-1894-z

Ji, G.-J., Liao, W., Chen, F.-F., Zhang, L., & Wang, K. (2017). Low-frequency blood oxygen level-dependent fluctuations in the brain white matter: More than just noise. *Science Bulletin*, *62*(9), 656–657. https://doi.org/10.1016/j.scib.2017.03.021

Jo, T., Nho, K., & Saykin, A. J. (2019). Deep Learning in Alzheimer's Disease: Diagnostic Classification and Prognostic Prediction Using Neuroimaging Data. *Frontiers in Aging Neuroscience*, *11*. https://doi.org/10.3389/fnagi.2019.00220

Jucker, M., & Walker, L. C. (2013). Self-propagation of pathogenic protein aggregates in neurodegenerative diseases. *Nature*, *501*(7465), 45–51. https://doi.org/10.1038/nature12481

Jungo, P., & Hewer, E. (2023). Code-free machine learning for classification of central nervous system histopathology images. *Journal of Neuropathology & Experimental Neurology*, nlac131. https://doi.org/10.1093/jnen/nlac131

Kaalund, S. S., Passamonti, L., Allinson, K. S. J., Murley, A. G., Robbins, T. W., Spillantini, M. G., & Rowe, J. B. (2020). Locus coeruleus pathology in progressive supranuclear

palsy, and its relation to disease severity. *Acta Neuropathologica Communications*, *8*(1), 11. https://doi.org/10.1186/s40478-020-0886-0

Kantara, C., O'Connell, M. R., Luthra, G., Gajjar, A., Sarkar, S., Ullrich, R. L., & Singh, P. (2015). Methods for detecting circulating cancer stem cells (CCSCs) as a novel approach for diagnosis of colon cancer relapse/metastasis. *Laboratory Investigation; a Journal of Technical Methods and Pathology*, *95*(1), 100–112. https://doi.org/10.1038/labinvest.2014.133

Kawamura, A., Abe, Y., Seki, F., Katayama, Y., Nishiyama, M., Takata, N., Tanaka, K. F., Okano, H., & Nakayama, K. I. (2020). Chd8 mutation in oligodendrocytes alters microstructure and functional connectivity in the mouse brain. *Molecular Brain*, *13*(1), 160. https://doi.org/10.1186/s13041-020-00699-x

Kayser, K., GÄśrtler, Jă., Bogovac, M., Bogovac, A., Goldmann, T., Vollmer, E., & Kayser, G. (2009). AI (artificial intelligence) in histopathology—From image analysis to automated diagnosis. *Folia Histochemica et Cytobiologica*, *47*(3), Article 3. https://doi.org/10.2478/v10042-009-0087-y

Khan, A. R., Ottoy, J., & Goubran, M. (2021). Towards a comprehensive 3D mapping of tau progression in early Alzheimer's disease. *Brain*, *144*(9), 2565–2567. https://doi.org/10.1093/brain/awab314

Kim, E.-J. (2020). Evidence of corticofugal tau spreading in patients with frontotemporal dementia. *Acta Neuropathologica*, 17.

Kim, K. H., Seo, J. D., Kim, E. S., Kim, H.-S., Jeon, S., Pak, K., Lee, M. J., Lee, J.-H., Lee, Y. M., Lee, K., Shin, J.-H., Ko, J. K., Jung, N.-Y., Lee, J. M., Yoon, J. A., Hwang, C., Ahn, J.-W., Sung, S., Spina, S., … Kim, E.-J. (2020). Comparison of Amyloid in Cerebrospinal Fluid, Brain Imaging, and Autopsy in a Case of Progressive

Supranuclear Palsy. *Alzheimer Disease & Associated Disorders*, *34*(3), 275–277. https://doi.org/10.1097/WAD.0000000000000396

Knopman, D. S., Parisi, J. E., Salviati, A., Floriach-Robert, M., Boeve, B. F., Ivnik, R. J., Smith, G. E., Dickson, D. W., Johnson, K. A., Petersen, L. E., McDonald, W. C., Braak, H., & Petersen, R. C. (2003). Neuropathology of cognitively normal elderly. *Journal of Neuropathology and Experimental Neurology*, *62*(11), 1087–1095. https://doi.org/10.1093/jnen/62.11.1087

Koga, S., Ghayal, N. B., & Dickson, D. W. (2021). Deep Learning-Based Image Classification in Differentiating Tufted Astrocytes, Astrocytic Plaques, and Neuritic Plaques. *Journal of Neuropathology & Experimental Neurology*, *80*(4), 306–312. https://doi.org/10.1093/jnen/nlab005

Koga, S., Ikeda, A., & Dickson, D. W. (2022). Deep learning-based model for diagnosing Alzheimer's disease and tauopathies. *Neuropathology and Applied Neurobiology*, *48*(1), e12759. https://doi.org/10.1111/nan.12759

Kong, Y., & Yu, T. (2018). A Deep Neural Network Model using Random Forest to Extract Feature Representation for Gene Expression Data Classification. *Scientific Reports*, *8*(1), Article 1. https://doi.org/10.1038/s41598-018-34833-6

Kovacs, G. G. (2015). Invited review: Neuropathology of tauopathies: principles and practice: Neuropathology of tauopathies. *Neuropathology and Applied Neurobiology*, *41*(1), 3–23. https://doi.org/10.1111/nan.12208

Kovacs, G. G. (2019). Molecular pathology of neurodegenerative diseases: Principles and practice. *Journal of Clinical Pathology*, *72*(11), 725–735. https://doi.org/10.1136/jclinpath-2019-205952

Kovacs, G. G., Alafuzoff, I., Al-Sarraj, S., Arzberger, T., Bogdanovic, N., Capellari, S., Ferrer, I., Gelpi, E., Kövari, V., Kretzschmar, H., Nagy, Z., Parchi, P., Seilhean, D.,

Soininen, H., Troakes, C., & Budka, H. (2008). Mixed Brain Pathologies in Dementia: The BrainNet Europe Consortium Experience. *Dementia and Geriatric Cognitive Disorders*, *26*(4), 343–350. https://doi.org/10.1159/000161560

Kovacs, G. G., Lukic, M. J., Irwin, D. J., Arzberger, T., Respondek, G., Lee, E. B., Coughlin, D., Giese, A., Grossman, M., Kurz, C., McMillan, C. T., Gelpi, E., Compta, Y., van Swieten, J. C., Laat, L. D., Troakes, C., Al-Sarraj, S., Robinson, J. L., Roeber, S., … Höglinger, G. U. (2020). Distribution patterns of tau pathology in progressive supranuclear palsy. *Acta Neuropathologica*, *140*(2), 99–119. https://doi.org/10.1007/s00401-020-02158-2

Kovacs, G. G., Milenkovic, I., Wöhrer, A., Höftberger, R., Gelpi, E., Haberler, C., Hönigschnabl, S., Reiner-Concin, A., Heinzl, H., Jungwirth, S., Krampla, W., Fischer, P., & Budka, H. (2013). Non-Alzheimer neurodegenerative pathologies and their combinations are more frequent than commonly believed in the elderly brain: A community-based autopsy series. *Acta Neuropathologica*, *126*(3), 365–384. https://doi.org/10.1007/s00401-013-1157-y

Kruschke, J. K. (2011). Bayesian Assessment of Null Values Via Parameter Estimation and Model Comparison. *Perspectives on Psychological Science*, *6*(3), 299–312. https://doi.org/10.1177/1745691611406925

Kruschke, J. K. (2018). Rejecting or Accepting Parameter Values in Bayesian Estimation. *Advances in Methods and Practices in Psychological Science*, *1*(2), 270–280. https://doi.org/10.1177/2515245918771304

Kruschke, J. K., Aguinis, H., & Joo, H. (2012). The Time Has Come: Bayesian Methods for Data Analysis in the Organizational Sciences. *Organizational Research Methods*, *15*(4), 722–752. https://doi.org/10.1177/1094428112457829

Kruschke, J. K., & Liddell, T. M. (2018a). Bayesian data analysis for newcomers. *Psychonomic Bulletin & Review*, *25*(1), 155–177. https://doi.org/10.3758/s13423-017-1272-1

Kruschke, J. K., & Liddell, T. M. (2018b). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*(1), 178–206. https://doi.org/10.3758/s13423-016-1221-4

La Joie, R., Perrotin, A., Barre, L., Hommet, C., Mezenge, F., Ibazizene, M., Camus, V., Abbas, A., Landeau, B., Guilloteau, D., de La Sayette, V., Eustache, F., Desgranges, B., & Chetelat, G. (2012). Region-Specific Hierarchy between Atrophy, Hypometabolism, and -Amyloid (A ) Load in Alzheimer's Disease Dementia. *Journal of Neuroscience*, *32*(46), 16265–16273. https://doi.org/10.1523/JNEUROSCI.2170-12.2012

Lee, H., Gu, Y., Seo, S. W., & Moon, S. H. (2023). Tau positron emission tomography in tauopathies: A narrative review. *Precision and Future Medicine*, *7*(1), 7–24. https://doi.org/10.23838/pfm.2023.00016

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research*, *18*(17), 1–5.

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

Ling, C. X., & Sheng, V. S. (2010). Cost-Sensitive Learning. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 231–235). Springer US. https://doi.org/10.1007/978-0-387-30164-8_181

Ling, H., de Silva, R., Massey, L. A., Courtney, R., Hondhamuni, G., Bajaj, N., Lowe, J., Holton, J. L., Lees, A., & Revesz, T. (2014). Characteristics of progressive supranuclear palsy presenting with corticobasal syndrome: A cortical variant. *Neuropathology and Applied Neurobiology*, *40*(2), 149–163. https://doi.org/10.1111/nan.12037

Lipton, Z. C., Elkan, C., & Naryanaswamy, B. (2014). Optimal Thresholding of Classifiers to Maximize F1 Measure. *Machine Learning and Knowledge Discovery in Databases : European Conference, ECML PKDD ... : Proceedings. ECML PKDD (Conference)*, *8725*, 225–239. https://doi.org/10.1007/978-3-662-44851-9_15

Litvan, I., Agid, Y., Calne, D., Campbell, G., Dubois, B., Duvoisin, R. C., Goetz, C. G., Golbe, L. I., Grafman, J., Growdon, J. H., Hallett, M., Jankovic, J., Quinn, N. P., Tolosa, E., & Zee, D. S. (1996). Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): Report of the NINDS-SPSP International Workshop. *Neurology*, *47*(1), 1–9. https://doi.org/10.1212/WNL.47.1.1

Lowe, V. J., Curran, G., Fang, P., Liesinger, A. M., Josephs, K. A., Parisi, J. E., Kantarci, K., Boeve, B. F., Pandey, M. K., Bruinsma, T., Knopman, D. S., Jones, D. T., Petrucelli, L., Cook, C. N., Graff-Radford, N. R., Dickson, D. W., Petersen, R. C., Jack, C. R., & Murray, M. E. (2016). An autoradiographic evaluation of AV-1451 Tau PET in dementia. *Acta Neuropathologica Communications*, *4*(1), 58. https://doi.org/10.1186/s40478-016-0315-6

Luo, H., Pan, X., Wang, Q., Ye, S., & Qian, Y. (2019). Logistic Regression and Random Forest for Effective Imbalanced Classification. *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, *1*, 916–917. https://doi.org/10.1109/COMPSAC.2019.00139

Marquié, M., Verwer, E. E., Meltzer, A. C., Kim, S. J. W., Agüero, C., Gonzalez, J., Makaretz, S. J., Siao Tick Chong, M., Ramanan, P., Amaral, A. C., Normandin, M. D., Vanderburg, C. R., Gomperts, S. N., Johnson, K. A., Frosch, M. P., & Gómez-Isla, T. (2017). Lessons learned about [F-18]-AV-1451 off-target binding from an autopsy-confirmed Parkinson's case. *Acta Neuropathologica Communications*, *5*. https://doi.org/10.1186/s40478-017-0482-0

Martinez-Lage, P., & Munoz, D. G. (1997). Prevalence and Disease Associations of Argyrophilic Grains of Braak. *Journal of Neuropathology & Experimental Neurology*, *56*(2), 157–164. https://doi.org/10.1097/00005072-199702000-00006

Marx, G. A., Koenigsberg, D. G., McKenzie, A. T., Kauffman, J., Hanson, R. W., Whitney, K., Signaevsky, M., Prastawa, M., Iida, M. A., White, C. L., Walker, J. M., Richardson, T. E., Koll, J., Fernandez, G., Zeineh, J., Cordon-Cardo, C., Crary, J. F., Farrell, K., & The PART working group. (2022). Artificial intelligence-derived neurofibrillary tangle burden is associated with antemortem cognitive impairment. *Acta Neuropathologica Communications*, *10*(1), 157. https://doi.org/10.1186/s40478-022-01457-x

MATLAB. (2010). *Version 7.10.0 (R2010a)*. The MathWorks Inc.

Mocanu, M.-M., Nissen, A., Eckermann, K., Khlistunova, I., Biernat, J., Drexler, D., Petrova, O., Schönig, K., Bujard, H., Mandelkow, E., Zhou, L., Rune, G., & Mandelkow, E.-M. (2008). The potential for beta-structure in the repeat domain of tau protein determines aggregation, synaptic decay, neuronal loss, and coassembly with endogenous Tau in inducible mouse models of tauopathy. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *28*(3), 737–748. https://doi.org/10.1523/JNEUROSCI.2824-07.2008

More, A. S., & Rana, D. P. (2017). Review of random forest classification techniques to resolve data imbalance. *2017 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, 72–78. https://doi.org/10.1109/ICISIM.2017.8122151

Murali, V. S., Chang, B.-J., Fiolka, R., Danuser, G., Cobanoglu, M. C., & Welf, E. S. (2019). An image-based assay to quantify changes in proliferation and viability upon drug treatment in 3D microenvironments. *BMC Cancer*, *19*(1), 502. https://doi.org/10.1186/s12885-019-5694-1

Narasimhan, S., Changolkar, L., Riddle, D. M., Kats, A., Stieber, A., Weitzman, S. A., Zhang, B., Li, Z., Roberson, E. D., Trojanowski, J. Q., & Lee, V. M. Y. (2019). Human tau pathology transmits glial tau aggregates in the absence of neuronal tau. *The Journal of Experimental Medicine*, *217*(2), e20190783. https://doi.org/10.1084/jem.20190783

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P. H., Albert, M., Boone, K., Miller, B. L., Cummings, J., & Benson, D. F. (1998). Frontotemporal lobar degeneration: A consensus on clinical diagnostic criteria. *Neurology*, *51*(6), 1546–1554. https://doi.org/10.1212/WNL.51.6.1546

Neltner, J. H., Abner, E. L., Schmitt, F. A., Denison, S. K., Anderson, S., Patel, E., & Nelson, P. T. (2012). Digital Pathology and Image Analysis for Robust High-Throughput Quantitative Assessment of Alzheimer Disease Neuropathologic Changes. *Journal of Neuropathology and Experimental Neurology*, *71*(12), 1075–1085. https://doi.org/10.1097/NEN.0b013e3182768de4

Oei, R. W., Hou, G., Liu, F., Zhong, J., Zhang, J., An, Z., Xu, L., & Yang, Y. (2019). Convolutional neural network for cell classification using microscope images of

intracellular actin networks. *PLoS ONE*, *14*(3), e0213626. https://doi.org/10.1371/journal.pone.0213626

Pan, P., Liu, Y., Zhang, Y., Zhao, H., Ye, X., & Xu, Y. (2017). Brain gray matter abnormalities in progressive supranuclear palsy revisited. *Oncotarget*, *8*(46), 80941–80955. https://doi.org/10.18632/oncotarget.20895

Pasquini, L., Nana, A. L., Toller, G., Brown, J. A., Deng, J., Staffaroni, A., Kim, E.-J., Hwang, J.-H. L., Li, L., Park, Y., Gaus, S. E., Allen, I., Sturm, V. E., Spina, S., Grinberg, L. T., Rankin, K. P., Kramer, J. H., Rosen, H. J., Miller, B. L., & Seeley, W. W. (2020). Salience Network Atrophy Links Neuron Type-Specific Pathobiology to Loss of Empathy in Frontotemporal Dementia. *Cerebral Cortex, 30*(10), 5387–5399. https://doi.org/10.1093/cercor/bhaa119

Pearson, K. (1901). LIII. *On lines and planes of closest fit to systems of points in space*. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Phan, H. T. H., Kumar, A., Kim, J., & Feng, D. (2016). Transfer learning of a convolutional neural network for HEp-2 cell image classification. *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, 1208–1211. https://doi.org/10.1109/ISBI.2016.7493483

Piattella, M. C., Upadhyay, N., Bologna, M., Sbardella, E., Tona, F., Formica, A., Petsas, N., Berardelli, A., & Pantano, P. (2015). Neuroimaging evidence of gray and white

matter damage and clinical correlates in progressive supranuclear palsy. *Journal of Neurology*, *262*(8), 1850–1858. https://doi.org/10.1007/s00415-015-7779-3

Plotly Technologies Inc. (2015). *Collaborative data science Publisher: Plotly Technologies Inc. Place of publication: Montréal, QC*. https://plot.ly

Popescu, V., Klaver, R., Voorn, P., Galis-de Graaf, Y., Knol, D., Twisk, J., Versteeg, A., Schenk, G., Van der Valk, P., Barkhof, F., De Vries, H., Vrenken, H., & Geurts, J. (2015). What drives MRI-measured cortical atrophy in multiple sclerosis? *Multiple Sclerosis Journal*, *21*(10), 1280–1290. https://doi.org/10.1177/1352458514562440

Prewitt, J. M. S., & Mendelsohn, M. L. (1966). The Analysis of Cell Images*. *Annals of the New York Academy of Sciences*, *128*(3), 1035–1053. https://doi.org/10.1111/j.1749-6632.1965.tb11715.x

Qian, K., Jiang, X., Liu, Z.-Q., Zhang, J., Fu, P., Su, Y., Brazhe, N. A., Liu, D., & Zhu, L.-Q. (2023). Revisiting the critical roles of reactive astrocytes in neurodegeneration. *Molecular Psychiatry*, 1–10. https://doi.org/10.1038/s41380-023-02061-8

Raamana, P. R. (2018). *VisualQC: Assistive tools for easy and rigorous quality control of neuroimaging data*. https://doi.org/10.5281/zenodo.1211365

Ravikumar, S., Wisse, L. E. M., Lim, S., Ittyerah, R., Xie, L., Bedard, M. L., Das, S. R., Lee, E. B., Tisdall, M. D., Prabhakaran, K., Lane, J., Detre, J. A., Mizsei, G., Trojanowski, J. Q., Robinson, J. L., Schuck, T., Grossman, M., Artacho-Pérula, E., de Onzoño Martin, M. M. I., … Yushkevich, P. A. (2021). Ex vivo MRI atlas of the human medial temporal lobe: Characterizing neurodegeneration due to tau pathology. *Acta Neuropathologica Communications*, *9*(1), 173. https://doi.org/10.1186/s40478-021-01275-7

Ribeiro, P. F. M., Ventura-Antunes, L., Gabi, M., Mota, B., Grinberg, L. T., Farfel, J. M., Ferretti-Rebustini, R. E. L., Leite, R. E. P., Filho, W. J., & Herculano-Houzel, S.

(2013). The human cerebral cortex is neither one nor many: Neuronal distribution reveals two quantitatively different zones in the gray matter, three in the white matter, and explains local variations in cortical folding. *Frontiers in Neuroanatomy*, *7*, 28. https://doi.org/10.3389/fnana.2013.00028

Rittman, T. (2020). Neurological update: Neuroimaging in dementia. *Journal of Neurology*. https://doi.org/10.1007/s00415-020-10040-0

Rittman, T., Rubinov, M., Vértes, P. E., Patel, A. X., Ginestet, C. E., Ghosh, B. C. P., Barker, R. A., Spillantini, M. G., Bullmore, E. T., & Rowe, J. B. (2016). Regional expression of the MAPT gene is associated with loss of hubs in brain networks and cognitive impairment in Parkinson disease and progressive supranuclear palsy. *Neurobiology of Aging*, *48*, 153–160. https://doi.org/10.1016/j.neurobiolaging.2016.09.001

Robinson, J. L., Yan, N., Caswell, C., Xie, S. X., Suh, E., Van Deerlin, V. M., Gibbons, G., Irwin, D. J., Grossman, M., Lee, E. B., Lee, V. M.-Y., Miller, B., & Trojanowski, J. Q. (2020). Primary Tau Pathology, Not Copathology, Correlates With Clinical Symptoms in PSP and CBD. *Journal of Neuropathology and Experimental Neurology*, *79*(3), 296–304. https://doi.org/10.1093/jnen/nlz141

Román, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., Amaducci, L., Orgogozo, J. M., Brun, A., & Hofman, A. (1993). Vascular dementia: Diagnostic criteria for research studies. Report of the NINDS-AIREN International Workshop. *Neurology*, *43*(2), 250–260. https://doi.org/10.1212/wnl.43.2.250

Rösler, T. W., Tayaranian Marvian, A., Brendel, M., Nykänen, N.-P., Höllerhage, M., Schwarz, S. C., Hopfner, F., Koeglsperger, T., Respondek, G., Schweyer, K., Levin, J., Villemagne, V. L., Barthel, H., Sabri, O., Müller, U., Meissner, W. G., Kovacs, G.

G., & Höglinger, G. U. (2019). Four-repeat tauopathies. *Progress in Neurobiology*, *180*, 101644. https://doi.org/10.1016/j.pneurobio.2019.101644

Salvesen, L., Winge, K., Brudek, T., Agander, T. K., Løkkegaard, A., & Pakkenberg, B. (2015). Neocortical Neuronal Loss in Patients with Multiple System Atrophy: A Stereological Study. *Cerebral Cortex*, bhv228. https://doi.org/10.1093/cercor/bhv228

Sander, K., Lashley, T., Gami, P., Gendron, T., Lythgoe, M. F., Rohrer, J. D., Schott, J. M., Revesz, T., Fox, N. C., & Årstad, E. (2016). Characterization of tau positron emission tomography tracer [18F]AV-1451 binding to postmortem tissue in Alzheimer's disease, primary tauopathies, and other dementias. *Alzheimer's & Dementia*, *12*(11), 1116–1124. https://doi.org/10.1016/j.jalz.2016.01.003

Scheltens, P., & Rockwood, K. (2011). How golden is the gold standard of neuropathology in dementia? *Alzheimer's & Dementia*, *7*(4), 486–489. https://doi.org/10.1016/j.jalz.2011.04.011

Schmidt, U., Weigert, M., Broaddus, C., & Myers, G. (2018). Cell Detection with Star-convex Polygons. *arXiv:1806.03535 [Cs]*, *11071*, 265–273. https://doi.org/10.1007/978-3-030-00934-2_30

Schott, J. M., Bartlett, J. W., Barnes, J., Leung, K. K., Ourselin, S., & Fox, N. C. (2010). Reduced sample sizes for atrophy outcomes in Alzheimer's disease trials: Baseline adjustment. *Neurobiology of Aging*, *31*(8), 1452-1462.e2. https://doi.org/10.1016/j.neurobiolaging.2010.04.011

Seeley, W. W., Crawford, R. K., Zhou, J., Miller, B. L., & Greicius, M. D. (2009). Neurodegenerative diseases target large-scale human brain networks. *Neuron*, *62*(1), 42–52. https://doi.org/10.1016/j.neuron.2009.03.024

Sexton, C., Snyder, H., Beher, D., Boxer, A. L., Brannelly, P., Brion, J.-P., Buée, L., Cacace, A. M., Chételat, G., Citron, M., DeVos, S. L., Diaz, K., Feldman, H. H., Frost, B.,

Goate, A. M., Gold, M., Hyman, B., Johnson, K., Karch, C. M., … Carrillo, M. C. (2022). Current directions in tau research: Highlights from Tau 2020. *Alzheimer's & Dementia, 18*(5), 988–1007. https://doi.org/10.1002/alz.12452

Shi, Y., Zhang, W., Yang, Y., Murzin, A. G., Falcon, B., Kotecha, A., van Beers, M., Tarutani, A., Kametani, F., Garringer, H. J., Vidal, R., Hallinan, G. I., Lashley, T., Saito, Y., Murayama, S., Yoshida, M., Tanaka, H., Kakita, A., Ikeuchi, T., … Scheres, S. H. W. (2021). Structure-based classification of tauopathies. *Nature*, *598*(7880), Article 7880. https://doi.org/10.1038/s41586-021-03911-7

Shifat-E-Rabbi, M., Yin, X., Fitzgerald, C. E., & Rohde, G. K. (2020). Cell Image Classification: A Comparative Overview. *Cytometry Part A*, *97*(4), 347–362. https://doi.org/10.1002/cyto.a.23984

Signaevsky, M., Prastawa, M., Farrell, K., Tabish, N., Baldwin, E., Han, N., Iida, M. A., Koll, J., Bryce, C., Purohit, D., Haroutunian, V., McKee, A. C., Stein, T. D., White, C. L., Walker, J., Richardson, T. E., Hanson, R., Donovan, M. J., Cordon-Cardo, C., … Crary, J. F. (2019). Artificial intelligence in neuropathology: Deep learning-based assessment of tauopathy. *Laboratory Investigation*, *99*(7), Article 7. https://doi.org/10.1038/s41374-019-0202-4

Solorzano, L., Wik, L., Bontell, T. O., Wang, Y., Klemm, A. H., Öfverstedt, J., Jakola, A. S., Östman, A., & Wählby, C. (2021). Machine learning for cell classification and neighborhood analysis in glioma tissue. *Cytometry Part A*, *n/a*(n/a). https://doi.org/10.1002/cyto.a.24467

Spillantini, M. G., & Goedert, M. (2013). Tau pathology and neurodegeneration. *The Lancet. Neurology*, *12*(6), 609–622. https://doi.org/10.1016/S1474-4422(13)70090-5

Spina, S., Brown, J. A., Deng, J., Gardner, R. C., Nana, A. L., Hwang, J.-H. L., Gaus, S. E., Huang, E. J., Kramer, J. H., Rosen, H. J., Kornak, J., Neuhaus, J., Miller, B. L.,

Grinberg, L. T., Boxer, A. L., & Seeley, W. W. (2019). Neuropathological correlates of structural and functional imaging biomarkers in 4-repeat tauopathies. *Brain*, *142*(7), 2068–2081. https://doi.org/10.1093/brain/awz122

Staffaroni, A. M., Elahi, F. M., McDermott, D., Marton, K., Karageorgiou, E., Sacco, S., Paoletti, M., Caverzasi, E., Hess, C. P., Rosen, H. J., & Geschwind, M. D. (2017). Neuroimaging in Dementia. *Seminars in Neurology*, *37*(5), 510–537. https://doi.org/10.1055/s-0037-1608808

Stezin, A., Lenka, A., Jhunjhunwala, K., Saini, J., & Pal, P. K. (2017). Advanced structural neuroimaging in progressive supranuclear palsy: Where do we stand? *Parkinsonism & Related Disorders*, *36*, 19–32. https://doi.org/10.1016/j.parkreldis.2016.12.023

Strimbu, K., & Tavel, J. A. (2010). What are Biomarkers? *Current Opinion in HIV and AIDS*, *5*(6), 463–466. https://doi.org/10.1097/COH.0b013e32833ed177

Tang, Z., Chuang, K. V., DeCarli, C., Jin, L.-W., Beckett, L., Keiser, M. J., & Dugger, B. N. (2019). Interpretable classification of Alzheimer's disease pathologies with a convolutional neural network pipeline. *Nature Communications*, *10*(1), 2173. https://doi.org/10.1038/s41467-019-10212-1

Tapiola, T., Pennanen, C., Tapiola, M., Tervo, S., Kivipelto, M., Hänninen, T., Pihlajamäki, M., Laakso, M. P., Hallikainen, M., Hämäläinen, A., Vanhanen, M., Helkala, E.-L., Vanninen, R., Nissinen, A., Rossi, R., Frisoni, G. B., & Soininen, H. (2008). MRI of hippocampus and entorhinal cortex in mild cognitive impairment: A follow-up study. *Neurobiology of Aging*, *29*(1), 31–38. https://doi.org/10.1016/j.neurobiolaging.2006.09.007

Togo, T., Sahara, N., Yen, S.-H., Cookson, N., Ishizawa, T., Hutton, M., de Silva, R., Lees, A., & Dickson, D. W. (2002). Argyrophilic Grain Disease Is a Sporadic 4-Repeat

Tauopathy. *Journal of Neuropathology & Experimental Neurology*, *61*(6), 547–556. https://doi.org/10.1093/jnen/61.6.547

Tsai, R. M., Lobach, I., Bang, J., Whitwell, J. L., Senjem, M. L., Jack, C. R., Rosen, H., Miller, B., & Boxer, A. L. (2016). Clinical correlates of longitudinal brain atrophy in progressive supranuclear palsy. *Parkinsonism & Related Disorders*, *28*, 29–35. https://doi.org/10.1016/j.parkreldis.2016.04.006

Vega, A. R., Chkheidze, R., Jarmale, V., Shang, P., Foong, C., Diamond, M. I., White, C. L., & Rajaram, S. (2021). Deep learning reveals disease-specific signatures of white matter pathology in tauopathies. *Acta Neuropathologica Communications*, *9*(1), 170. https://doi.org/10.1186/s40478-021-01271-x

Vemuri, P., Simon, G., Kantarci, K., Whitwell, J. L., Senjem, M. L., Przybelski, S. A., Gunter, J. L., Josephs, K. A., Knopman, D. S., Boeve, B. F., Ferman, T. J., Dickson, D. W., Parisi, J. E., Petersen, R. C., & Jack, C. R. (2011). Antemortem differential diagnosis of dementia pathology using structural MRI: Differential-STAND. *NeuroImage*, *55*(2), 522–531. https://doi.org/10.1016/j.neuroimage.2010.12.073

Vemuri, P., Wiste, H. J., Weigand, S. D., Shaw, L. M., Trojanowski, J. Q., Weiner, M. W., Knopman, D. S., Petersen, R. C., & Jack, C. R. (2009). MRI and CSF biomarkers in normal, MCI, and AD subjects. *Neurology*, *73*(4), 294–301. https://doi.org/10.1212/WNL.0b013e3181af79fb

Visser, P. J., Verhey, F. R. J., Hofman, P. a. M., Scheltens, P., & Jolles, J. (2002). Medial temporal lobe atrophy predicts Alzheimer's disease in patients with minor cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry*, *72*(4), 491–497. https://doi.org/10.1136/jnnp.72.4.491

Vizcarra, J. C., Gearing, M., Keiser, M. J., Glass, J. D., Dugger, B. N., & Gutman, D. A. (2020). Validation of machine learning models to detect amyloid pathologies across

institutions. *Acta Neuropathologica Communications*, *8*(1), 59. https://doi.org/10.1186/s40478-020-00927-4

Vogel, J. W., Young, A. L., Oxtoby, N. P., Smith, R., Ossenkoppele, R., Strandberg, O. T., La Joie, R., Aksman, L. M., Grothe, M. J., Iturria-Medina, Y., Pontecorvo, M. J., Devous, M. D., Rabinovici, G. D., Alexander, D. C., Lyoo, C. H., Evans, A. C., & Hansson, O. (2021). Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nature Medicine*, *27*(5), 871–881. https://doi.org/10.1038/s41591-021-01309-6

von Bartheld, C. S., Bahney, J., & Herculano-Houzel, S. (2016). The Search for True Numbers of Neurons and Glial Cells in the Human Brain: A Review of 150 Years of Cell Counting. *The Journal of Comparative Neurology*, *524*(18), 3865–3895. https://doi.org/10.1002/cne.24040

Walker, L., McAleese, K. E., Johnson, M., Khundakar, A. A., Erskine, D., Thomas, A. J., McKeith, I. G., & Attems, J. (2017). Quantitative neuropathology: An update on automated methodologies and implications for large scale cohorts. *Journal of Neural Transmission (Vienna, Austria: 1996)*, *124*(6), 671–683. https://doi.org/10.1007/s00702-017-1702-2

WallÃ¸e, S., Pakkenberg, B., & Fabricius, K. (2014). Stereological estimation of total cell numbers in the human cerebral and cerebellar cortex. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00508

Wang, Y., & Mandelkow, E. (2016). Tau in physiology and pathology. *Nature Reviews Neuroscience*, *17*(1), Article 1. https://doi.org/10.1038/nrn.2015.1

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

West, M. J., & Gundersen, H. J. (1990). Unbiased stereological estimation of the number of neurons in the human hippocampus. *The Journal of Comparative Neurology*, *296*(1), 1–22. https://doi.org/10.1002/cne.902960102

Whiteside, D. J., Jones, P. S., Ghosh, B. C. P., Coyle-Gilchrist, I., Gerhard, A., Hu, M. T., Klein, J. C., Leigh, P. N., Church, A., Burn, D. J., Morris, H. R., Rowe, J. B., & Rittman, T. (2021). Altered network stability in progressive supranuclear palsy. *Neurobiology of Aging*, *107*, 109–117. https://doi.org/10.1016/j.neurobiolaging.2021.07.007

Whitwell, J. L., Dickson, D. W., Murray, M. E., Weigand, S. D., Tosakulwong, N., Senjem, M. L., Knopman, D. S., Boeve, B. F., Parisi, J. E., Petersen, R. C., Jack, C. R., & Josephs, K. A. (2012). Neuroimaging correlates of pathologically defined subtypes of Alzheimer's disease: A case-control study. *The Lancet. Neurology*, *11*(10), 868–877. https://doi.org/10.1016/S1474-4422(12)70200-4

Whitwell, J. L., Josephs, K. A., Murray, M. E., Kantarci, K., Przybelski, S. A., Weigand, S. D., Vemuri, P., Senjem, M. L., Parisi, J. E., Knopman, D. S., Boeve, B. F., Petersen, R. C., Dickson, D. W., & Jack, C. R. (2008). MRI correlates of neurofibrillary tangle pathology at autopsy: A voxel-based morphometry study. *Neurology*, *71*(10), 743–749. https://doi.org/10.1212/01.wnl.0000324924.91351.7d

Williams, D. R., Holton, J. L., Strand, C., Pittman, A., de Silva, R., Lees, A. J., & Revesz, T. (2007). Pathological tau burden and distribution distinguishes progressive supranuclear palsy-parkinsonism from Richardson's syndrome. *Brain*, *130*(6), 1566–1576. https://doi.org/10.1093/brain/awm104

Wittmann, C. W., Wszolek, M. F., Shulman, J. M., Salvaterra, P. M., Lewis, J., Hutton, M., & Feany, M. B. (2001). Tauopathy in Drosophila: Neurodegeneration Without

Neurofibrillary Tangles. *Science*, *293*(5530), 711–714. https://doi.org/10.1126/science.1062382

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

World Health Organization. (2023, March 15). *Dementia*. https://www.who.int/news-room/fact-sheets/detail/dementia

Xu, J. J., Henstock, P. V., Dunn, M. C., Smith, A. R., Chabot, J. R., & de Graaf, D. (2008). Cellular imaging predictions of clinical drug-induced liver injury. *Toxicological Sciences: An Official Journal of the Society of Toxicology*, *105*(1), 97–105. https://doi.org/10.1093/toxsci/kfn109

Yushkevich, P. A., Muñoz López, M., Iñiguez de Onzoño Martin, M. M., Ittyerah, R., Lim, S., Ravikumar, S., Bedard, M. L., Pickup, S., Liu, W., Wang, J., Hung, L. Y., Lasserve, J., Vergnet, N., Xie, L., Dong, M., Cui, S., McCollum, L., Robinson, J. L., Schuck, T., … Insausti, R. (2021). Three-dimensional mapping of neurofibrillary tangle burden in the human medial temporal lobe. *Brain: A Journal of Neurology*, *144*(9), 2784–2797. https://doi.org/10.1093/brain/awab262

Zhang, X., Gweon, H., & Provost, S. (2020). Threshold Moving Approaches for Addressing the Class Imbalance Problem and their Application to Multi-label Classification. *2020 4th International Conference on Advances in Image Processing*, 72–77. https://doi.org/10.1145/3441250.3441274

Zhang, Y., Wu, K.-M., Yang, L., Dong, Q., & Yu, J.-T. (2022). Tauopathies: New perspectives and challenges. *Molecular Neurodegeneration*, *17*(1), 28. https://doi.org/10.1186/s13024-022-00533-z

Zhou, J., Gennatas, E. D., Kramer, J. H., Miller, B. L., & Seeley, W. W. (2012). Predicting regional neurodegeneration from the healthy brain functional connectome. *Neuron*, *73*(6), 1216–1227. https://doi.org/10.1016/j.neuron.2012.03.004

Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Finding the Best Classification Threshold in Imbalanced Classification. *Big Data Research*, *5*, 2–8. https://doi.org/10.1016/j.bdr.2015.12.001

# Appendix

In Appendix, extra information is provided for analyses carried out in chapter 5 and 6. For chapter 5, model diagnostic plots are provided Bayesian regression models when assessing the relationship between neuropathological severity (PSP stage, tau burden) and clinical severity (PSPRS score). For chapter 6, full results for w-score, weighted degree (WD) and closeness centrality (CC) in PSP are provided for 258 regions in the analysis (where applicable). Due to a large number of models constructed when assessing pathology-imaging relationship, here, I have provided model diagnostic results for only key results, 1) atrophy-tau relationship in dentate nucleus and 2) weighted degree and tau at whole brain level.

# Model fit for chapter 5

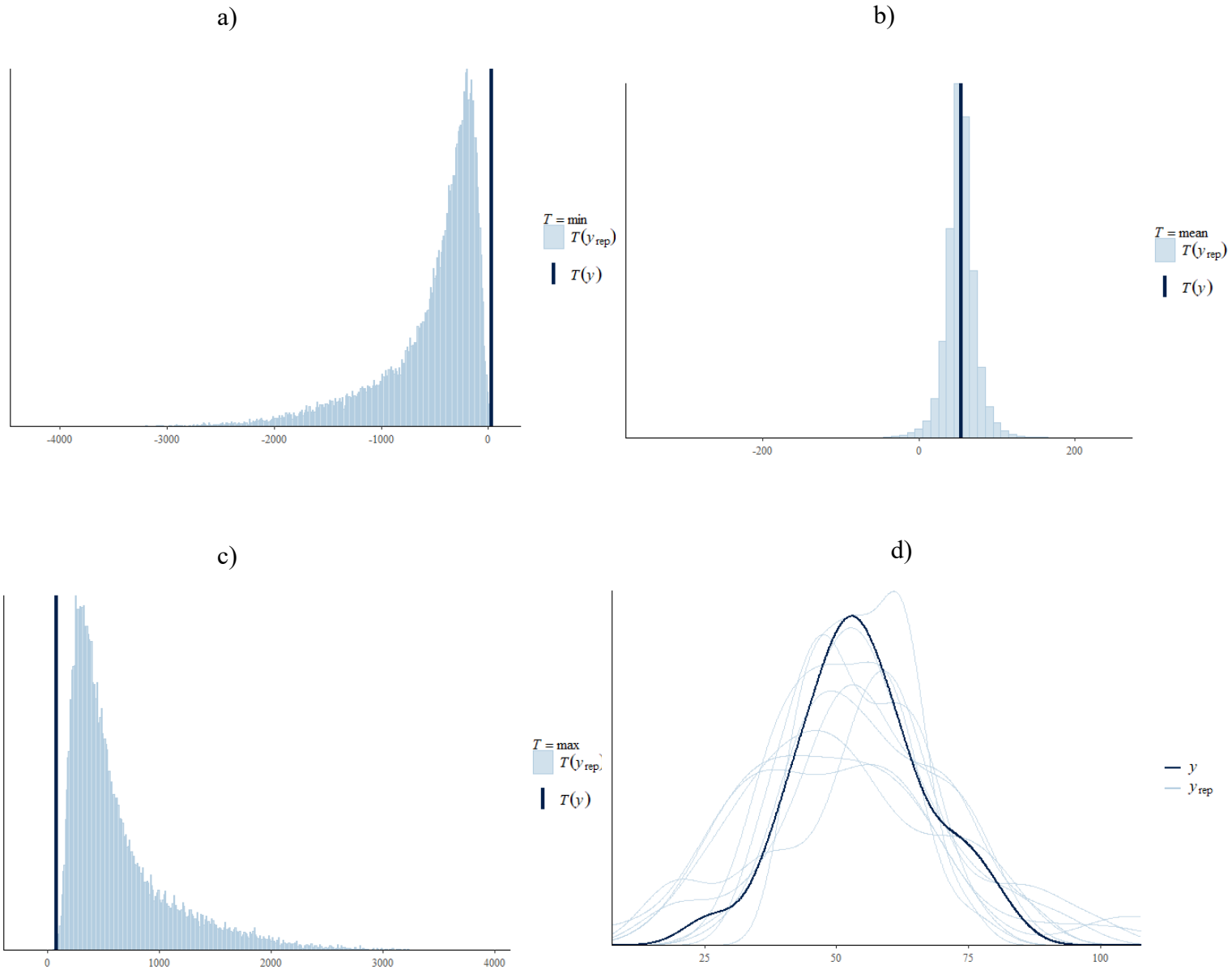



Figure 0.1: Prior predictive check of N(0,100) for the model, PSPRS score ~ PSP stage + disease duration + PSPRS-death interval. T(y) is the distribution of the data (dark blue line), T($y_{rep}$) is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

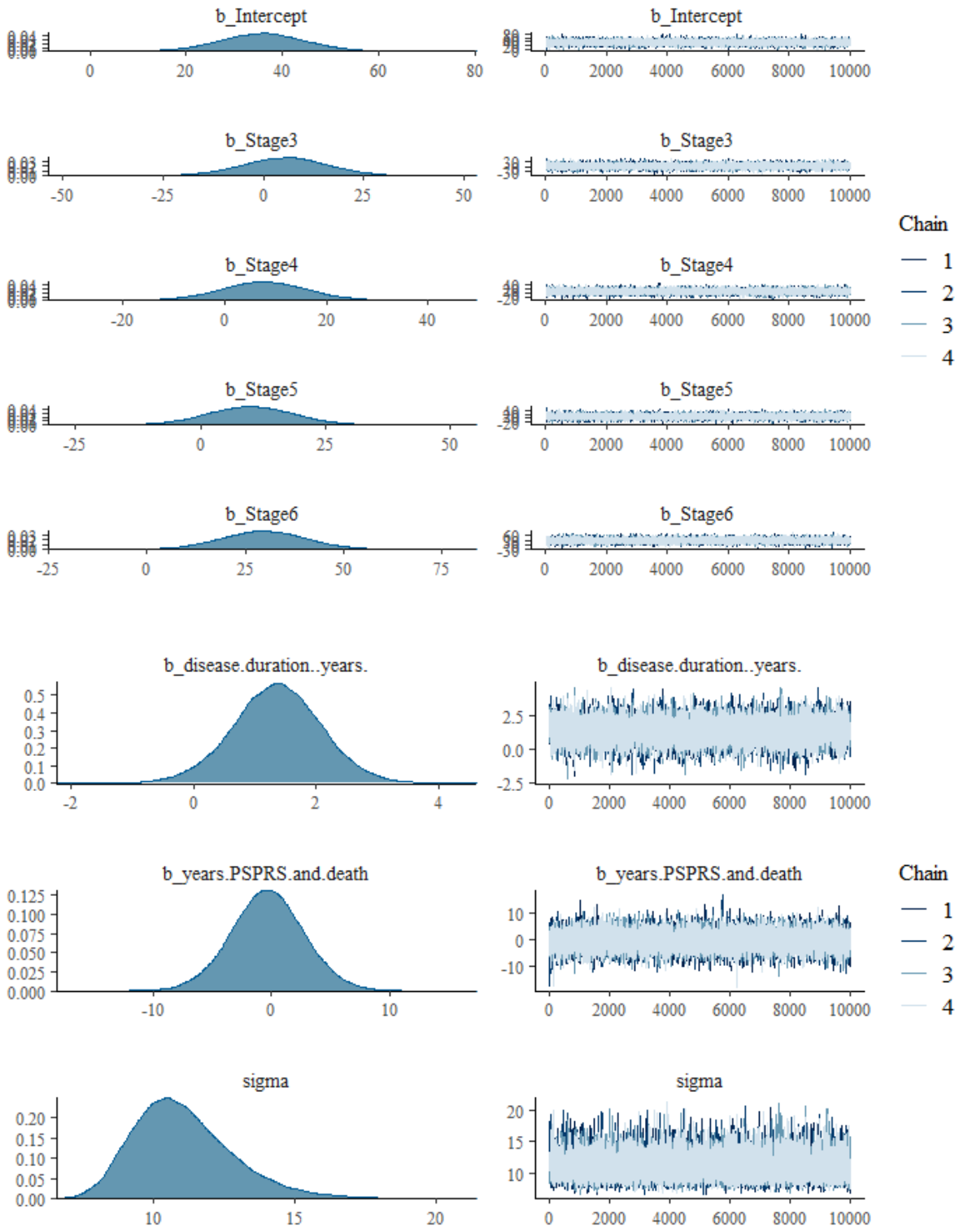



Figure 0.2: Plots of posterior distribution (left) and trace plots (right) of the regression coefficients in the model, PSPRS score ~ PSP stage + disease duration + PSPRS-death interval. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.

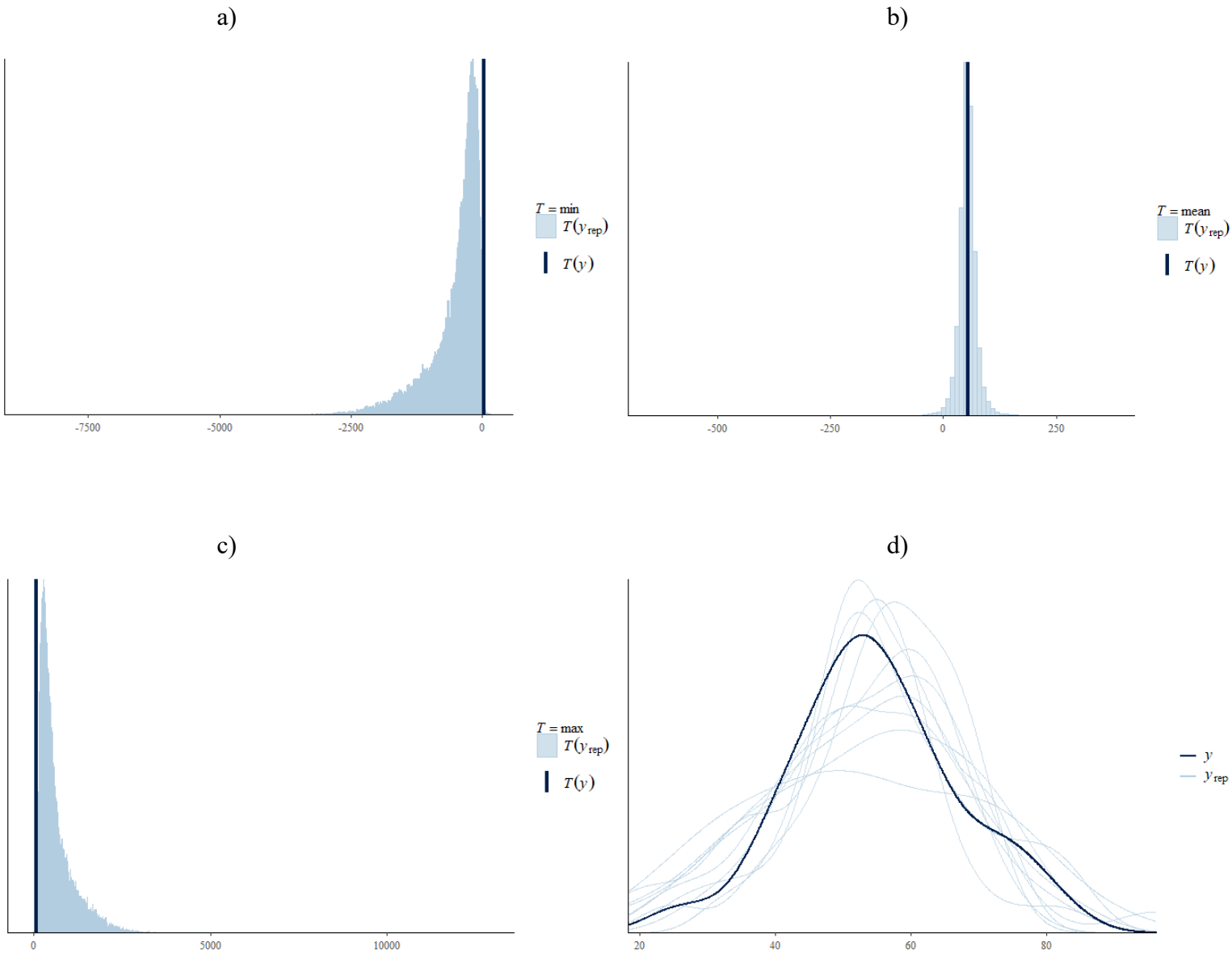



Figure 0.3: Prior predictive check of N(0,100) for the model, PSPRS score ~ total tau burden (all regions) + disease duration + PSPRS-death interval. T(y) is the distribution of the data (dark blue line), $T(y_{rep})$ is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

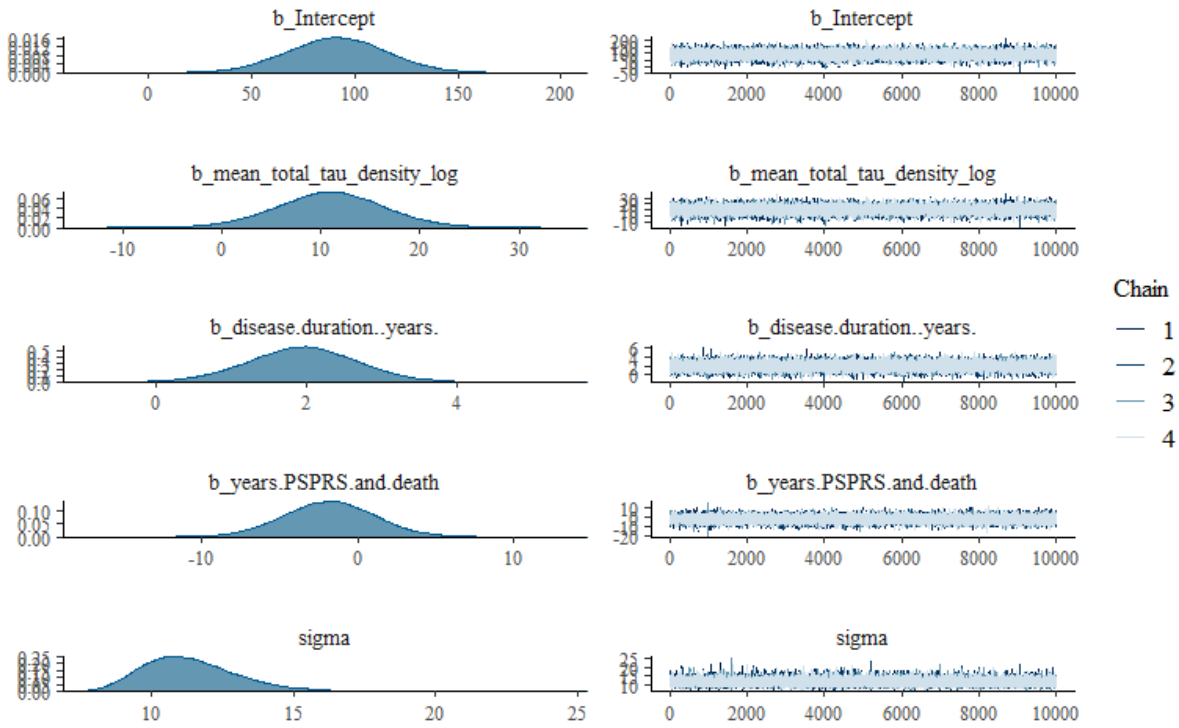



Figure 0.4: Plots of posterior distribution and trace plots of the regression coefficient in the model; PSPRS score ~ total tau burden (all regions) + disease duration + PSPRS-death interval. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.

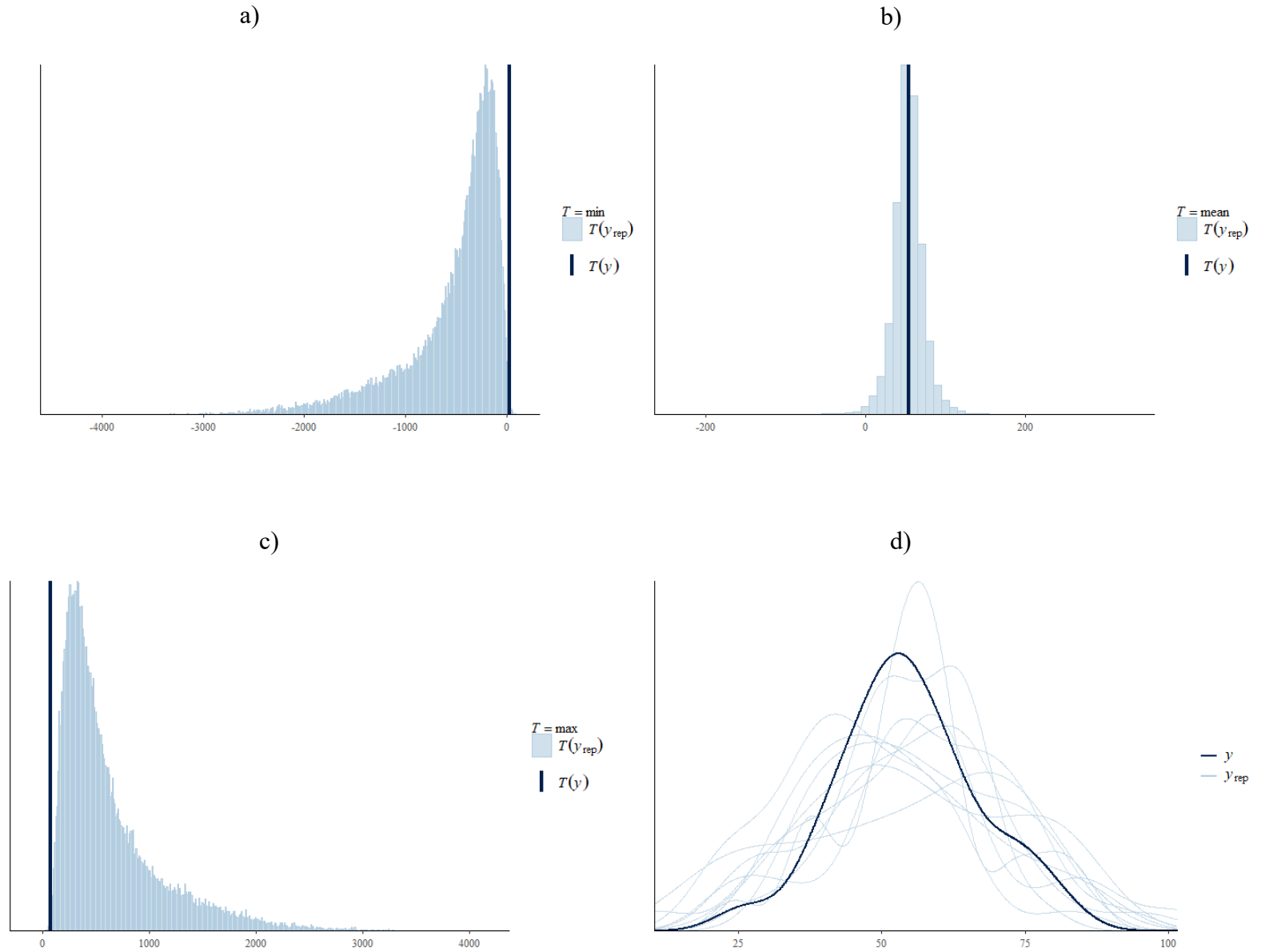



Figure 0.5: Prior predictive check of N(0,100) for the model, PSPRS score ~ total tau burden (cortical regions) + disease duration + PSPRS-death interval. T(y) is the distribution of the data (dark blue line), $T(y_{rep})$ is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

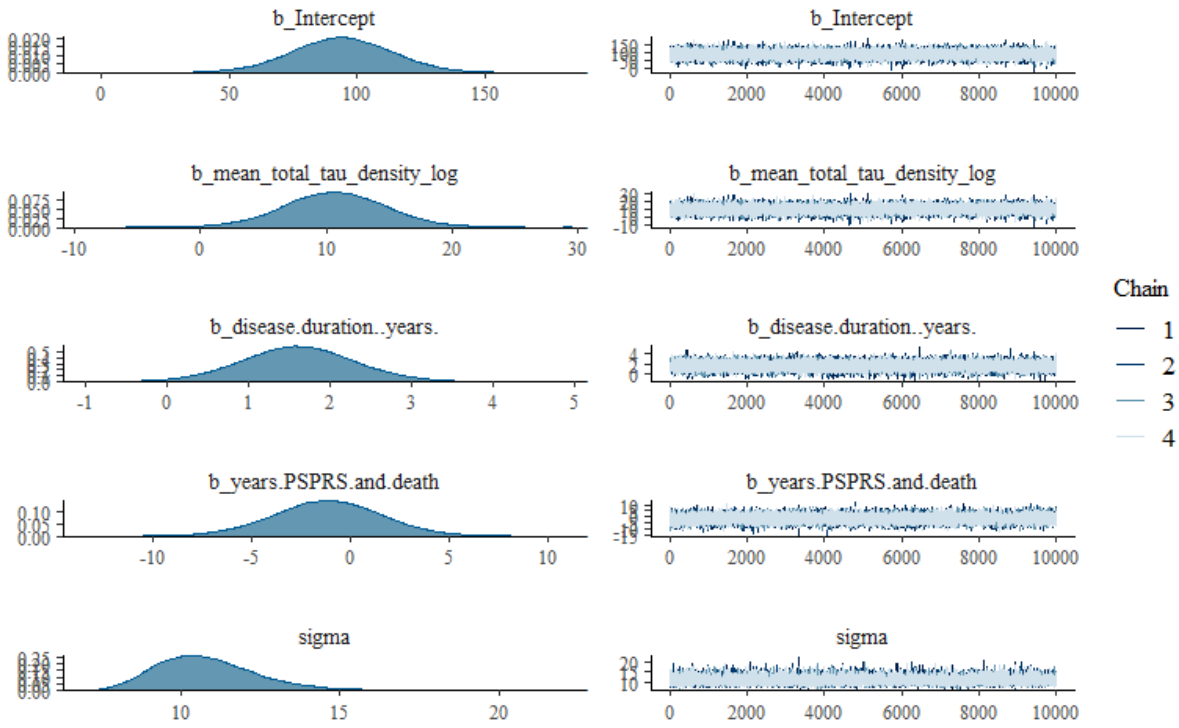



Figure 0.6: Plots of posterior distribution and trace plots of the regression coefficient in the model; PSPRS score ~ total tau burden (cortical regions) + disease duration + PSPRS-death interval. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.

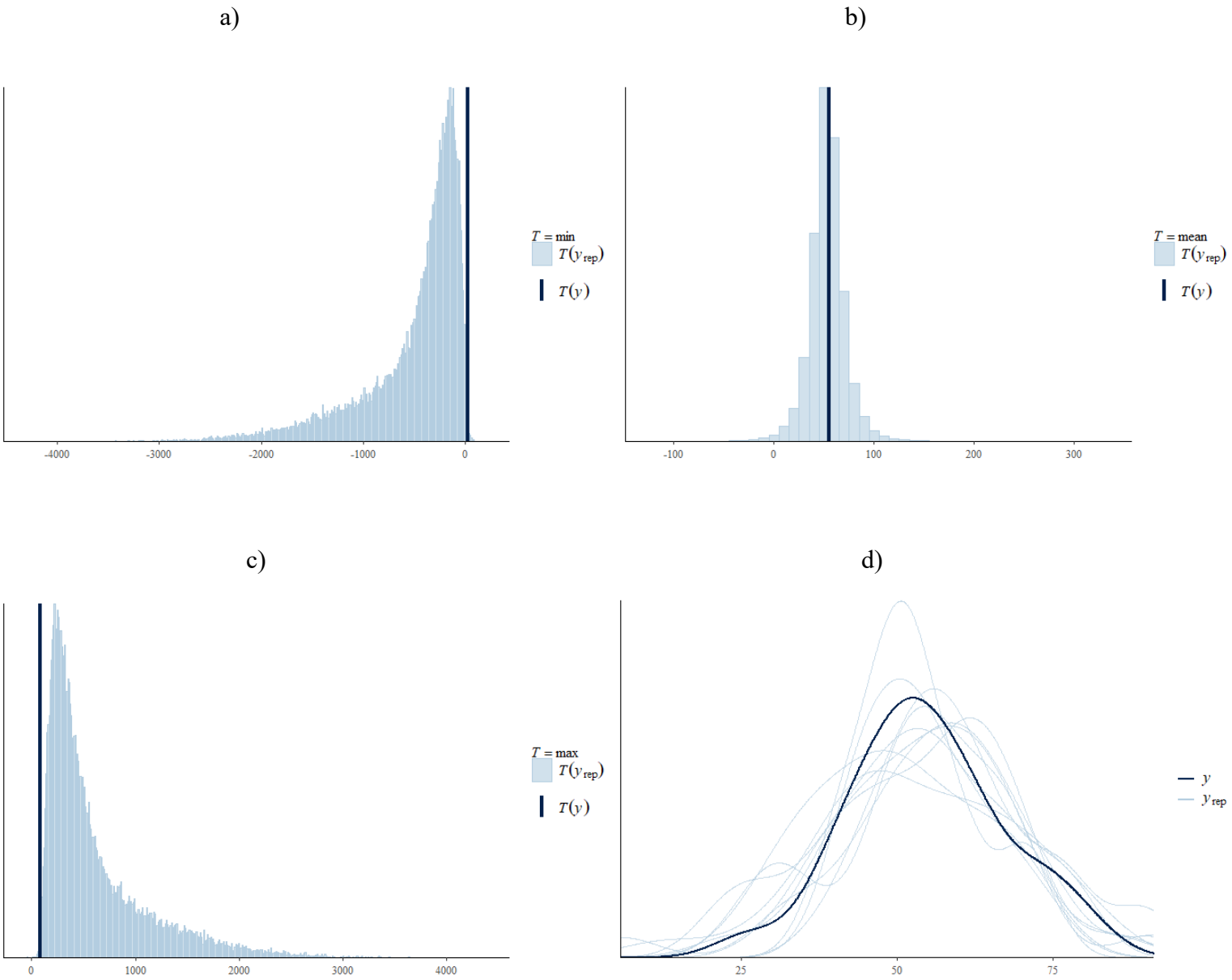



Figure 0.7: Prior predictive check of N(0,100) for the model, PSPRS score ~ neurofibrillary tangle burden (subcortical regions) + disease duration + PSPRS-death interval. T(y) is the distribution of the data (dark blue line), $T(y_{rep})$ is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

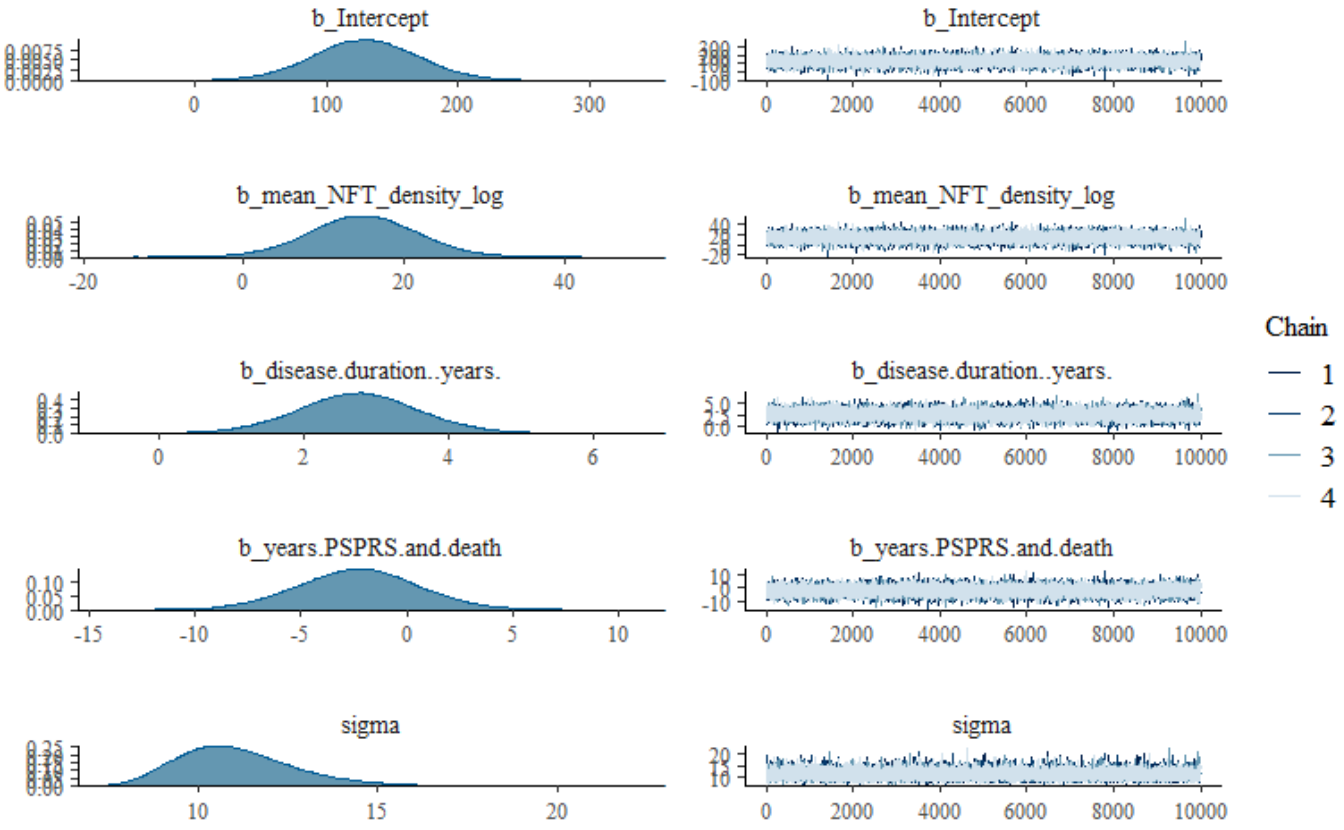



Figure 0.8: Plots of posterior distribution and trace plots of the regression coefficient in the model; PSPRS score ~ neurofibrillary tangle burden (subcortical regions) + disease duration + PSPRS-death interval. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.

# Analysis results & model fit for chapter 6

Table 0.1: W-score, weighted degree (WD) and closeness centrality (CC) are reported for all 258 brain regions in the analysis. Not applicable (NA) is reported where no calculation was done.

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 1 | Amygdala | medial amyg | mAmyg_L | 0.282 | 13.516 | 0.773 |
| 2 | Amygdala | medial amyg | mAmyg_R | 0.228 | 13.504 | 0.766 |
| 3 | Amygdala | lateral amyg | lAmyg_L | 0.884 | 13.254 | 0.781 |
| 4 | Amygdala | lateral amyg | lAmyg_R | 0.867 | 13.796 | 0.762 |
| 5 | Hippocampus | rostral hipp | rHipp_L | 0.461 | 17.506 | 0.769 |
| 6 | Hippocampus | rostral hipp | rHipp_R | 0.320 | 16.170 | 0.765 |
| 7 | Hippocampus | caudal hipp | cHipp_L | 0.576 | 14.637 | 0.772 |
| 8 | Hippocampus | caudal hipp | cHipp_R | 0.919 | 14.967 | 0.776 |
| 9 | Basal Ganglia | ventral caudate | vCa_L | 1.554 | 12.440 | 0.755 |
| 10 | Basal Ganglia | ventral caudate | vCa_R | 1.328 | 14.339 | 0.758 |
| 11 | Basal Ganglia | globus pallidus | GP_L | 1.317 | 15.068 | 0.742 |
| 12 | Basal Ganglia | globus pallidus | GP_R | 1.392 | 15.984 | 0.750 |
| 13 | Basal Ganglia | nucleus accumbens | NAC_L | 0.705 | 15.224 | 0.772 |
| 14 | Basal Ganglia | nucleus accumbens | NAC_R | 1.428 | 14.815 | 0.764 |
| 15 | Basal Ganglia | ventromedial putamen | vmPu_L | 1.820 | 15.600 | 0.750 |
| 16 | Basal Ganglia | ventromedial putamen | vmPu_R | 1.300 | 15.059 | 0.759 |
| 17 | Basal Ganglia | dorsal caudate | dCa_L | 0.681 | 11.834 | 0.743 |
| 18 | Basal Ganglia | dorsal caudate | dCa_R | 0.713 | 12.544 | 0.740 |
| 19 | Basal Ganglia | dorsolateral putamen | dlPu_L | 1.977 | 15.957 | 0.761 |
| 20 | Basal Ganglia | dorsolateral putamen | dlPu_R | 1.565 | 15.459 | 0.766 |
| 21 | Thalamus | medial pre-frontal thalamus | mPFtha_L | 1.565 | 19.729 | 0.737 |
| 22 | Thalamus | medial pre-frontal thalamus | mPFtha_R | 1.417 | 19.526 | 0.732 |
| 23 | Thalamus | pre-motor thalamus | mPMtha_L | 0.538 | 14.075 | 0.736 |
| 24 | Thalamus | pre-motor thalamus | mPMtha_R | 1.544 | 18.136 | 0.742 |
| 25 | Thalamus | sensory thalamus | Stha_L | 1.699 | 16.568 | 0.756 |
| 26 | Thalamus | sensory thalamus | Stha_R | 1.550 | 16.030 | 0.765 |
| 27 | Thalamus | rostral temporal thalamus | rTtha_L | -0.153 | 17.006 | 0.742 |
| 28 | Thalamus | rostral temporal thalamus | rTtha_R | 0.044 | 15.741 | 0.728 |
| 29 | Thalamus | posterior parietal thalamus | PPtha_L | 1.818 | 19.998 | 0.748 |
| 30 | Thalamus | posterior parietal thalamus | PPtha_R | 1.845 | 17.460 | 0.766 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 31 | Thalamus | occipital thalamus | Otha_L | 1.516 | 17.275 | 0.763 |
| 32 | Thalamus | occipital thalamus | Otha_R | 1.014 | 14.322 | 0.756 |
| 33 | Thalamus | caudal temporal thalamus | cTtha_L | 1.128 | 14.068 | 0.755 |
| 34 | Thalamus | caudal temporal thalamus | cTtha_R | -0.022 | 12.664 | 0.752 |
| 35 | Thalamus | lateral pre-frontal thalamus | lPFtha_L | 1.904 | 20.166 | 0.738 |
| 36 | Thalamus | lateral pre-frontal thalamus | lPFtha_R | 1.797 | 21.280 | 0.743 |
| 37 | Subthalamic nucleus | Subthalamic nucleus | STN_L | 3.901 | NA | NA |
| 38 | Subthalamic nucleus | Subthalamic nucleus | STN_R | 3.407 | NA | NA |
| 39 | Superior frontal gyrus | medial area 8 | A8m_L | 0.852 | 17.418 | 0.769 |
| 40 | Superior frontal gyrus | medial area 9 | A8m_R | 0.330 | 18.085 | 0.777 |
| 41 | Superior frontal gyrus | dorsolateral area 8 | A8dl_L | 0.350 | 16.230 | 0.740 |
| 42 | Superior frontal gyrus | dorsolateral area 9 | A8dl_R | 0.336 | 16.001 | 0.742 |
| 43 | Superior frontal gyrus | lateral area 9 | A9l_L | 0.250 | 14.537 | 0.733 |
| 44 | Superior frontal gyrus | lateral area 10 | A9l_R | -0.219 | 15.390 | 0.739 |
| 45 | Superior frontal gyrus | dorsolateral area 6 | A6dl_L | 0.448 | 16.619 | 0.769 |
| 46 | Superior frontal gyrus | dorsolateral area 7 | A6dl_R | 0.470 | 15.811 | 0.768 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 47 | Superior frontal gyrus | medial area 6 | A6m_L | 1.116 | 21.085 | 0.763 |
| 48 | Superior frontal gyrus | medial area 7 | A6m_R | 0.877 | 22.147 | 0.758 |
| 49 | Superior frontal gyrus | medial area 9 | A9m_L | 0.829 | 18.564 | 0.748 |
| 50 | Superior frontal gyrus | medial area 9 | A9m_R | 0.509 | 18.807 | 0.754 |
| 51 | Superior frontal gyrus | medial area 10 | A10m_L | 0.702 | 17.155 | 0.735 |
| 52 | Superior frontal gyrus | medial area 11 | A10m_R | -0.081 | 17.509 | 0.735 |
| 53 | Middle frontal gyrus | dorsal area 9/46 | A9/46d_L | 0.464 | 15.599 | 0.771 |
| 54 | Middle frontal gyrus | dorsal area 9/46 | A9/46d_R | 0.263 | 16.871 | 0.762 |
| 55 | Middle frontal gyrus | inferior frontal junction | IFJ_L | 0.730 | 16.790 | 0.757 |
| 56 | Middle frontal gyrus | inferior frontal junction | IFJ_R | 0.770 | 17.064 | 0.743 |
| 57 | Middle frontal gyrus | area46 | A46_L | 0.651 | 15.635 | 0.762 |
| 58 | Middle frontal gyrus | area46 | A46_R | 0.078 | 16.120 | 0.754 |
| 59 | Middle frontal gyrus | ventral area 9/46 | A9/46v_R | 0.546 | 16.075 | 0.743 |
| 60 | Middle frontal gyrus | ventral area 9/46 | A9/46v_R | 0.030 | 16.075 | 0.743 |
| 61 | Middle frontal gyrus | ventral lateral area 8 | A8vl_L | 0.426 | 16.318 | 0.740 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 62 | Middle frontal gyrus | ventral lateral area 8 | A8vl_R | 0.418 | 15.236 | 0.742 |
| 63 | Middle frontal gyrus | ventral lateral area 6 | A6vl_L | 0.793 | 16.704 | 0.768 |
| 64 | Middle frontal gyrus | ventral lateral area 6 | A6vl_R | 0.132 | 16.845 | 0.758 |
| 65 | Middle frontal gyrus | lateral area 10 | A10l_L | 0.325 | 14.604 | 0.757 |
| 66 | Middle frontal gyrus | lateral area 10 | A10l_R | -0.108 | 15.093 | 0.755 |
| 67 | Inferior frontal gyrus | dorsal area44 | A44d_L | 0.568 | 15.439 | 0.762 |
| 68 | Inferior frontal gyrus | dorsal area44 | A44d_R | 0.636 | 16.139 | 0.755 |
| 69 | Inferior frontal gyrus | inferior frontal sulcus | IFS_L | 0.505 | 14.228 | 0.761 |
| 70 | Inferior frontal gyrus | inferior frontal sulcus | IFS_R | 0.448 | 14.635 | 0.757 |
| 71 | Inferior frontal gyrus | caudal area 45 | A45c_L | 0.186 | 12.734 | 0.754 |
| 72 | Inferior frontal gyrus | caudal area 45 | A45c_R | 0.428 | 14.661 | 0.748 |
| 73 | Inferior frontal gyrus | rostral area 45 | A45r_L | 0.400 | 13.713 | 0.761 |
| 74 | Inferior frontal gyrus | rostral area 45 | A45r_R | 0.058 | 13.757 | 0.754 |
| 75 | Inferior frontal gyrus | opercular area 44 | A44op_L | 0.654 | 16.192 | 0.763 |
| 76 | Inferior frontal gyrus | opercular area 44 | A44op_R | 0.553 | 16.593 | 0.764 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 77 | Inferior frontal gyrus | ventral area 44 | A44v_L | 0.530 | 14.748 | 0.757 |
| 78 | Inferior frontal gyrus | ventral area 45 | A44v_R | 0.642 | 15.864 | 0.753 |
| 79 | Orbital gyrus | medial area 14 | A14m_L | 0.151 | 15.607 | 0.731 |
| 80 | Orbital gyrus | medial area 14 | A14m_R | 0.023 | 16.698 | 0.729 |
| 81 | Orbital gyrus | orbital area 14/47 | A12/47o_L | 0.673 | 15.076 | 0.770 |
| 82 | Orbital gyrus | orbital area 14/47 | A12/47o_R | 0.185 | 14.312 | 0.760 |
| 83 | Orbital gyrus | lateral area 11 | A11l_L | 0.319 | 14.188 | 0.774 |
| 84 | Orbital gyrus | lateral area 11 | A11l_R | 0.133 | 15.437 | 0.764 |
| 85 | Orbital gyrus | medial area 11 | A11m_L | 0.342 | 14.000 | 0.736 |
| 86 | Orbital gyrus | medial area 11 | A11m_R | -0.406 | 14.572 | 0.737 |
| 87 | Orbital gyrus | area 13 | A13_L | 0.471 | 13.926 | 0.763 |
| 88 | Orbital gyrus | area 13 | A13_R | 0.256 | 13.668 | 0.761 |
| 89 | Orbital gyrus | lateral area 12/47 | A12/47l_L | 0.126 | 15.662 | 0.770 |
| 90 | Orbital gyrus | lateral area 12/47 | A12/47l_R | 0.496 | 16.630 | 0.766 |
| 91 | Precentral gyrus | area 4 (head&face region) | A4hf_L | 1.484 | 19.989 | 0.754 |
| 92 | Precentral gyrus | area 4 (head&face region) | A4hf_R | 0.865 | 17.964 | 0.746 |
| 93 | Precentral gyrus | caudal dorsolateral area 6 | A6cdl_L | 1.037 | 20.173 | 0.758 |
| 94 | Precentral gyrus | caudal dorsolateral area 6 | A6cdl_R | 0.860 | 20.329 | 0.762 |
| 95 | Precentral gyrus | area 4 (upper limb region) | A4ul_L | 0.891 | 22.661 | 0.738 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 96 | Precentral gyrus | area 4 (upper limb region) | A4ul_R | 0.708 | 20.967 | 0.737 |
| 97 | Precentral gyrus | area 4 (trunk reggion) | A4t_L | 1.134 | 18.200 | 0.748 |
| 98 | Precentral gyrus | area 4 (trunk reggion) | A4t_R | 0.975 | 19.492 | 0.746 |
| 99 | Precentral gyrus | area 4 (tongue & larynx) | A4tl_L | 0.521 | 19.361 | 0.745 |
| 100 | Precentral gyrus | area 4 (tongue & larynx) | A4tl_R | 0.671 | 19.030 | 0.746 |
| 101 | Precentral gyrus | caudal ventrolateral area 6 | A6cvl_L | 1.014 | 17.611 | 0.762 |
| 102 | Precentral gyrus | caudal ventrolateral area 6 | A6cvl_R | 0.630 | 18.922 | 0.755 |
| 103 | Paracentral lobule | area1/2/3 (lower limb region) | A1/2/3ll_L | 0.039 | 22.138 | 0.760 |
| 104 | Paracentral lobule | area1/2/3 (lower limb region) | A1/2/3ll_R | 0.387 | 25.051 | 0.751 |
| 105 | Paracentral lobule | area 4 (lower limb region) | A4ll_L | 0.884 | 22.060 | 0.744 |
| 106 | Paracentral lobule | area 4 (lower limb region) | A4ll_R | 0.978 | 23.336 | 0.743 |
| 107 | Superior temporal gyrus | medial area 38 | A38m_L | -0.293 | 15.579 | 0.759 |
| 108 | Superior temporal gyrus | medial area 38 | A38m_R | -0.531 | 16.214 | 0.762 |
| 109 | Superior temporal gyrus | area 41/42 | A41/42_L | 0.647 | 21.561 | 0.739 |
| 110 | Superior temporal gyrus | area 41/42 | A41/42_R | 0.423 | 21.321 | 0.744 |
| 111 | Superior temporal gyrus | TE1.0 and TE1.2_L | TE1.0 and TE1.2_L | 0.611 | 23.543 | 0.750 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 111 | Superior temporal gyrus | TE1.0 and TE1.2_L | TE1.0 and TE1.2_L | 0.611 | 23.543 | 0.750 |
| 112 | Superior temporal gyrus | TE1.0 and TE1.2_L | TE1.0 and TE1.2_R | 0.458 | 22.357 | 0.748 |
| 113 | Superior temporal gyrus | caudal area 22 | A22c_L | 1.052 | 18.914 | 0.756 |
| 114 | Superior temporal gyrus | caudal area 22 | A22c_R | 0.465 | 18.778 | 0.757 |
| 115 | Superior temporal gyrus | lateral area 38 | A38l_L | 0.692 | 14.887 | 0.779 |
| 116 | Superior temporal gyrus | lateral area 38 | A38l_R | 0.515 | 16.739 | 0.772 |
| 117 | Superior temporal gyrus | rostral area 22 | A22r_L | 0.815 | 19.176 | 0.761 |
| 118 | Superior temporal gyrus | rostral area 22 | A22r_R | 0.365 | 19.261 | 0.774 |
| 119 | Middle temporal gyrus | caudal area 21 | A21c_L | 0.723 | 14.907 | 0.745 |
| 120 | Middle temporal gyrus | caudal area 21 | A21c_R | 0.327 | 14.263 | 0.747 |
| 121 | Middle temporal gyrus | rostral area 21 | A21r_L | 0.366 | 14.098 | 0.755 |
| 122 | Middle temporal gyrus | rostral area 21 | A21r_R | 0.013 | 15.461 | 0.755 |
| 123 | Middle temporal gyrus | dorsolateral area 37 | A37dl_L | 0.152 | 16.269 | 0.765 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 124 | Middle temporal gyrus | dorsolateral area 37 | A37dl_R | 0.400 | 16.210 | 0.763 |
| 125 | Middle temporal gyrus | anteiror superior temporal sulcus | aSTS_L | 0.394 | 16.552 | 0.756 |
| 126 | Middle temporal gyrus | anteiror superior temporal sulcus | aSTS_R | 0.436 | 18.164 | 0.763 |
| 127 | Inferior temporal gyrus | intermediate ventral area 20 | A20iv_L | 0.665 | 12.560 | 0.760 |
| 128 | Inferior temporal gyrus | intermediate ventral area 20 | A20iv_R | 0.586 | 12.358 | 0.753 |
| 129 | Inferior temporal gyrus | extreme lateroventral area 37 | A37elv_L | 0.085 | 15.820 | 0.766 |
| 130 | Inferior temporal gyrus | extreme lateroventral area 37 | A37elv_R | 0.448 | 15.377 | 0.766 |
| 131 | Inferior temporal gyrus | rostral area 20 | A20r_L | 0.023 | 12.226 | 0.752 |
| 132 | Inferior temporal gyrus | rostral area 20 | A20r_R | -0.055 | 12.383 | 0.755 |
| 133 | Inferior temporal gyrus | intermediate lateral area 20 | A20il_L | 0.767 | 13.296 | 0.744 |
| 134 | Inferior temporal gyrus | intermediate lateral area 20 | A20il_R | 0.409 | 12.999 | 0.737 |
| 135 | Inferior temporal gyrus | ventrolateral area 37 | A37vl_L | 0.489 | 14.383 | 0.761 |
| 136 | Inferior temporal gyrus | ventrolateral area 37 | A37vl_R | 0.678 | 15.464 | 0.760 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 137 | Inferior temporal gyrus | caudolateral of area 20 | A20cl_L | 0.027 | 12.773 | 0.740 |
| 138 | Inferior temporal gyrus | caudolateral of area 20 | A20cl_R | 0.372 | 12.313 | 0.749 |
| 139 | Inferior temporal gyrus | caudoventral of area 20 | A20cv_L | 0.450 | 11.458 | 0.749 |
| 140 | Inferior temporal gyrus | caudoventral of area 20 | A20cv_R | 0.330 | 11.530 | 0.752 |
| 141 | Fusiform gyrus | rostroventral area 20 | A20rv_L | 0.683 | 15.987 | 0.761 |
| 142 | Fusiform gyrus | rostroventral area 20 | A20rv_R | 0.750 | 16.914 | 0.764 |
| 143 | Fusiform gyrus | medioventral area 37 | A37mv_L | 0.821 | 21.909 | 0.739 |
| 144 | Fusiform gyrus | medioventral area 37 | A37mv_R | 0.719 | 21.755 | 0.740 |
| 145 | Fusiform gyrus | ventrolatral area37 | A37lv_L | 0.751 | 17.692 | 0.766 |
| 146 | Fusiform gyrus | ventrolatral area37 | A37lv_R | 0.478 | 19.185 | 0.756 |
| 147 | Parahippocampal gyrus | rostral area 35/36 | A35/36r_L | 0.396 | 13.126 | 0.756 |
| 148 | Parahippocampal gyrus | rostral area 35/36 | A35/36r_R | 0.275 | 13.955 | 0.748 |
| 149 | Parahippocampal gyrus | caudal area 35/36 | A35/36c_L | 0.399 | 13.168 | 0.771 |
| 150 | Parahippocampal gyrus | caudal area 35/36 | A35/36c_R | 0.433 | 13.024 | 0.765 |
| 151 | Parahippocampal gyrus | lateral posterior parahippocampal gyrus (PPHC) | TL_L | 0.160 | 13.862 | 0.769 |
| 152 | Parahippocampal gyrus | lateral posterior parahippocampal gyrus (PPHC) | TL_R | 0.571 | 14.133 | 0.768 |
| 153 | Parahippocampal gyrus | area 28/34 | A28/34_L | 0.673 | 14.794 | 0.763 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 154 | Parahippocampal gyrus | area 28/34 | A28/34_R | -0.210 | 13.589 | 0.754 |
| 155 | Parahippocampal gyrus | temporal agranular insular cortex | TI_L | 0.015 | 11.849 | 0.756 |
| 156 | Parahippocampal gyrus | temporal agranular insular cortex | TI_R | 0.391 | 11.574 | 0.750 |
| 157 | Parahippocampal gyrus | medial PPHC | TH_L | 0.028 | 15.029 | 0.780 |
| 158 | Parahippocampal gyrus | medial PPHC | TH_R | -0.122 | 15.893 | 0.782 |
| 159 | Posterior superior temporal sulcus | rostroposterior superior temporal sulcus (rpSTS) | rpSTS_L | 0.755 | 15.696 | 0.767 |
| 160 | Posterior superior temporal sulcus | rostroposterior superior temporal sulcus (rpSTS) | rpSTS_R | 0.638 | 16.987 | 0.765 |
| 161 | Posterior superior temporal sulcus | caudoposterior superior temporal sulcus (cpSTS) | cpSTS_L | 0.684 | 16.086 | 0.768 |
| 162 | Posterior superior temporal sulcus | caudoposterior superior temporal sulcus (cpSTS) | cpSTS_R | 0.516 | 16.673 | 0.762 |
| 163 | Superior parietal lobule | rostral area 7 | A7r_L | 0.098 | 16.967 | 0.750 |
| 164 | Superior parietal lobule | rostral area 7 | A7r_R | 0.409 | 17.418 | 0.731 |
| 165 | Superior parietal lobule | caudal area 7 | A7c_L | 0.474 | 16.970 | 0.745 |
| 166 | Superior parietal lobule | caudal area 7 | A7c_R | 0.613 | 16.288 | 0.744 |
| 167 | Superior parietal lobule | lateral area 5 | A5l_L | 0.336 | 20.303 | 0.748 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 168 | Superior parietal lobule | lateral area 5 | A5l_R | 0.325 | 17.660 | 0.747 |
| 169 | Superior parietal lobule | postcentral area 7 | A7pc_L | 0.041 | 18.136 | 0.749 |
| 170 | Superior parietal lobule | postcentral area 7 | A7pc_R | 0.228 | 17.813 | 0.738 |
| 171 | Superior parietal lobule | intraparietal area 7 | A7ip_L | 0.319 | 19.312 | 0.752 |
| 172 | Superior parietal lobule | intraparietal area 7 | A7ip_R | 0.705 | 18.745 | 0.740 |
| 173 | Inferior parietal lobule | caudal area 39 | A39c_L | 0.443 | 16.233 | 0.751 |
| 174 | Inferior parietal lobule | caudal area 39 | A39c_R | 0.431 | 16.331 | 0.762 |
| 175 | Inferior parietal lobule | rostrodorsal area 39 | A39rd_L | 0.201 | 15.599 | 0.743 |
| 176 | Inferior parietal lobule | rostrodorsal area 39 | A39rd_R | 0.434 | 15.370 | 0.739 |
| 177 | Inferior parietal lobule | rostrodorsal area 40 | A40rd_L | 0.151 | 19.155 | 0.755 |
| 178 | Inferior parietal lobule | rostrodorsal area 40 | A40rd_R | 0.588 | 18.841 | 0.750 |
| 179 | Inferior parietal lobule | caudal area 40 | A40c_L | 0.342 | 15.036 | 0.758 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 180 | Inferior parietal lobule | caudal area 40 | A40c_R | 0.446 | 13.970 | 0.745 |
| 181 | Inferior parietal lobule | rostroventral area 39 | A39rv_L | 0.581 | 17.558 | 0.742 |
| 182 | Inferior parietal lobule | rostroventral area 39 | A39rv_R | 0.314 | 16.446 | 0.742 |
| 183 | Inferior parietal lobule | rostralventral area 40 | A40rv_L | 0.023 | 18.850 | 0.746 |
| 184 | Inferior parietal lobule | rostralventral area 40 | A40rv_R | 0.564 | 18.671 | 0.744 |
| 185 | Precuneus | medial area 7 (PEp) | A7m_L | -0.157 | 18.333 | 0.752 |
| 186 | Precuneus | medial area 7 (PEp) | A7m_R | 0.465 | 17.977 | 0.750 |
| 187 | Precuneus | medial area 5 (PEm) | A5m_L | 0.428 | 22.261 | 0.763 |
| 188 | Precuneus | medial area 5 (PEm) | A5m_R | 0.624 | 24.388 | 0.755 |
| 189 | Precuneus | dorsomedial parietooccipital sulcus (PEr) | dmPOS_L | 0.615 | 19.128 | 0.752 |
| 190 | Precuneus | dorsomedial parietooccipital sulcus (PEr) | dmPOS_R | 0.839 | 19.232 | 0.755 |
| 191 | Precuneus | area 31 (LC1) | A31_L | 0.341 | 19.781 | 0.738 |
| 192 | Precuneus | area 31 (LC1) | A31_R | 0.720 | 20.638 | 0.734 |
| 193 | Postcentral gyrus | area 1/2/3 (upper limb, head & face region) | A1/2/3ulhf_L | 0.618 | 19.819 | 0.734 |
| 194 | Postcentral gyrus | area 1/2/3 (upper limb, head & face region) | A1/2/3ulhf_R | 0.735 | 20.379 | 0.739 |
| 195 | Postcentral gyrus | area 1/2/3 (tongue & larynx region) | A1/2/3tonIa_L | 0.463 | 19.436 | 0.745 |
| 196 | Postcentral gyrus | area 1/2/3 (tongue & larynx region) | A1/2/3tonIa_R | 0.659 | 19.875 | 0.741 |
| 197 | Postcentral gyrus | area 2 | A2_L | 0.519 | 20.456 | 0.737 |
| 198 | Postcentral gyrus | area 2 | A2_R | 0.702 | 20.339 | 0.739 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 199 | Postcentral gyrus | area 1/2/3 (trunk region) | A1/2/3tru_L | 0.246 | 21.674 | 0.744 |
| 200 | Postcentral gyrus | area 1/2/3 (trunk region) | A1/2/3tru_R | 0.154 | 22.860 | 0.740 |
| 201 | Insula gyrus | hypergranular insular | G_L | 0.251 | 18.411 | 0.761 |
| 202 | Insula gyrus | hypergranular insular | G_R | 0.605 | 16.534 | 0.759 |
| 203 | Insula gyrus | ventral agranular insular | vIa_L | 0.008 | 13.324 | 0.780 |
| 204 | Insula gyrus | ventral agranular insular | vIa_R | 0.114 | 13.925 | 0.774 |
| 205 | Insula gyrus | dorsal agranular insular | dIa_L | 0.159 | 15.933 | 0.752 |
| 206 | Insula gyrus | dorsal agranular insular | dIa_R | 0.850 | 15.497 | 0.749 |
| 207 | Insula gyrus | ventral granular insular | vId/vIg_L | -0.076 | 16.722 | 0.759 |
| 208 | Insula gyrus | ventral granular insular | vId/vIg_R | -0.124 | 17.802 | 0.750 |
| 209 | Insula gyrus | dorsal granular insular | dIg_L | 0.738 | 17.039 | 0.756 |
| 210 | Insula gyrus | dorsal granular insular | dIg_R | 0.473 | 17.311 | 0.755 |
| 211 | Insula gyrus | dorsal dyspranular insular | dId_L | 0.379 | 17.069 | 0.732 |
| 212 | Insula gyrus | dorsal dyspranular insular | dId_R | 0.581 | 17.545 | 0.739 |
| 213 | Cingulate gyrus | dorsal area 23 | A23d_L | 0.169 | 20.365 | 0.729 |
| 214 | Cingulate gyrus | dorsal area 23 | A23d_R | 0.395 | 19.493 | 0.742 |
| 215 | Cingulate gyrus | rostroventral area 24 | A24rv_L | 0.137 | 19.488 | 0.776 |
| 216 | Cingulate gyrus | rostroventral area 24 | A24rv_R | 0.319 | 16.886 | 0.765 |
| 217 | Cingulate gyrus | pregenual area 32 | A32p_L | 0.250 | 17.137 | 0.761 |
| 218 | Cingulate gyrus | pregenual area 32 | A32p_R | 0.055 | 17.866 | 0.766 |
| 219 | Cingulate gyrus | ventral area 23 | A23v_L | 0.159 | 19.049 | 0.747 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 220 | Cingulate gyrus | ventral area 23 | A23v_R | 0.148 | 19.701 | 0.747 |
| 221 | Cingulate gyrus | caudodorsal area 24 | A24cd_L | 0.298 | 21.087 | 0.764 |
| 222 | Cingulate gyrus | caudodorsal area 24 | A24cd_R | 0.473 | 21.318 | 0.754 |
| 223 | Cingulate gyrus | caudal area 23 | A23c_L | 0.506 | 22.299 | 0.785 |
| 224 | Cingulate gyrus | caudal area 23 | A23c_R | 0.323 | 23.678 | 0.770 |
| 225 | Cingulate gyrus | subgenual area 32 | A32sg_L | 0.336 | 15.023 | 0.737 |
| 226 | Cingulate gyrus | subgenual area 32 | A32sg_R | 0.215 | 15.542 | 0.748 |
| 227 | MedioVentral Occipital cortex | caudal lingual gyrus | cLinG_L | 0.404 | 20.823 | 0.729 |
| 228 | MedioVentral Occipital cortex | caudal lingual gyrus | cLinG_R | 0.054 | 19.803 | 0.730 |
| 229 | MedioVentral Occipital cortex | rostral cuneus gyrus | rCunG_L | 0.298 | 21.370 | 0.724 |
| 230 | MedioVentral Occipital cortex | rostral cuneus gyrus | rCunG_R | 0.262 | 21.318 | 0.732 |
| 231 | MedioVentral Occipital cortex | caudal cuneus gyrus | cCunG_L | 0.049 | 18.522 | 0.725 |
| 232 | MedioVentral Occipital cortex | caudal cuneus gyrus | cCunG_R | -0.022 | 19.105 | 0.718 |
| 233 | MedioVentral Occipital cortex | rostral lingual gyrus | rLinG_L | 0.181 | 21.171 | 0.736 |
| 234 | MedioVentral Occipital cortex | rostral lingual gyrus | rLinG_R | 0.143 | 21.435 | 0.733 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 235 | MedioVentral Occipital cortex | ventomedial parietoocipital sulcus | vmPOS_L | 0.392 | 21.435 | 0.743 |
| 236 | MedioVentral Occipital cortex | ventomedial parietoocipital sulcus | vmPOS_R | 0.082 | 21.232 | 0.744 |
| 237 | Lateral occipital cortex | middle occipital gyrus | mOccG_L | 0.106 | 17.648 | 0.713 |
| 238 | Lateral occipital cortex | middle occipital gyrus | mOccG_R | 0.070 | 17.449 | 0.718 |
| 239 | Lateral occipital cortex | area V5/MT+ | V5/MT+_L | 0.320 | 18.607 | 0.732 |
| 240 | Lateral occipital cortex | area V5/MT+ | V5/MT+_R | 0.468 | 17.758 | 0.731 |
| 241 | Lateral occipital cortex | occipital polar cortex | OPC_L | -0.099 | 17.685 | 0.719 |
| 242 | Lateral occipital cortex | occipital polar cortex | OPC_R | 0.191 | 16.967 | 0.726 |
| 243 | Lateral occipital cortex | inferior occipital gyrus | iOccG_L | -0.078 | 18.103 | 0.723 |
| 244 | Lateral occipital cortex | inferior occipital gyrus | iOccG_R | -0.029 | 18.777 | 0.728 |
| 245 | Lateral occipital cortex | medial superior occipital gyrus | msOccG_L | 0.236 | 18.530 | 0.717 |
| 246 | Lateral occipital cortex | medial superior occipital gyrus | msOccG_R | 0.549 | 18.463 | 0.722 |
| 247 | Lateral occipital cortex | lateral superior occipital gyrus | lsOccG_L | 0.426 | 18.090 | 0.753 |

| Index | Region | Specific region | Parcel | w-score | WD | CC |
|---|---|---|---|---|---|---|
| 248 | Lateral occipital cortex | lateral superior occipital gyrus | lsOccG_R | 0.484 | 18.816 | 0.747 |
| 249 | Cerebellum | Cerebellum white matter | Left.Cerebellum.White.Matter | 1.301 | NA | NA |
| 250 | Cerebellum | Cerebellum cortex | Left.Cerebellum.Cortex | 0.780 | NA | NA |
| 251 | Cerebellum | Cerebellum white matter | Right.Cerebellum.White.Matter | 1.678 | NA | NA |
| 252 | Cerebellum | Cerebellum cortex | Right.Cerebellum.Cortex | 0.526 | NA | NA |
| 253 | Midbrain | Midbrain | Midbrain | 3.095 | NA | NA |
| 254 | Brainstem | Pons | Pons | 1.684 | NA | NA |
| 255 | Brainstem | Medulla | Medulla | 1.367 | NA | NA |
| 256 | Brainstem | Superior cerebellar peduncle | SCP | 1.108 | NA | NA |
| 258 | Cerebellum | Dentate nucleus | DN_L | 1.548 | NA | NA |
| 258 | Cerebellum | Dentate nucleus | DN_R | 1.491 | NA | NA |

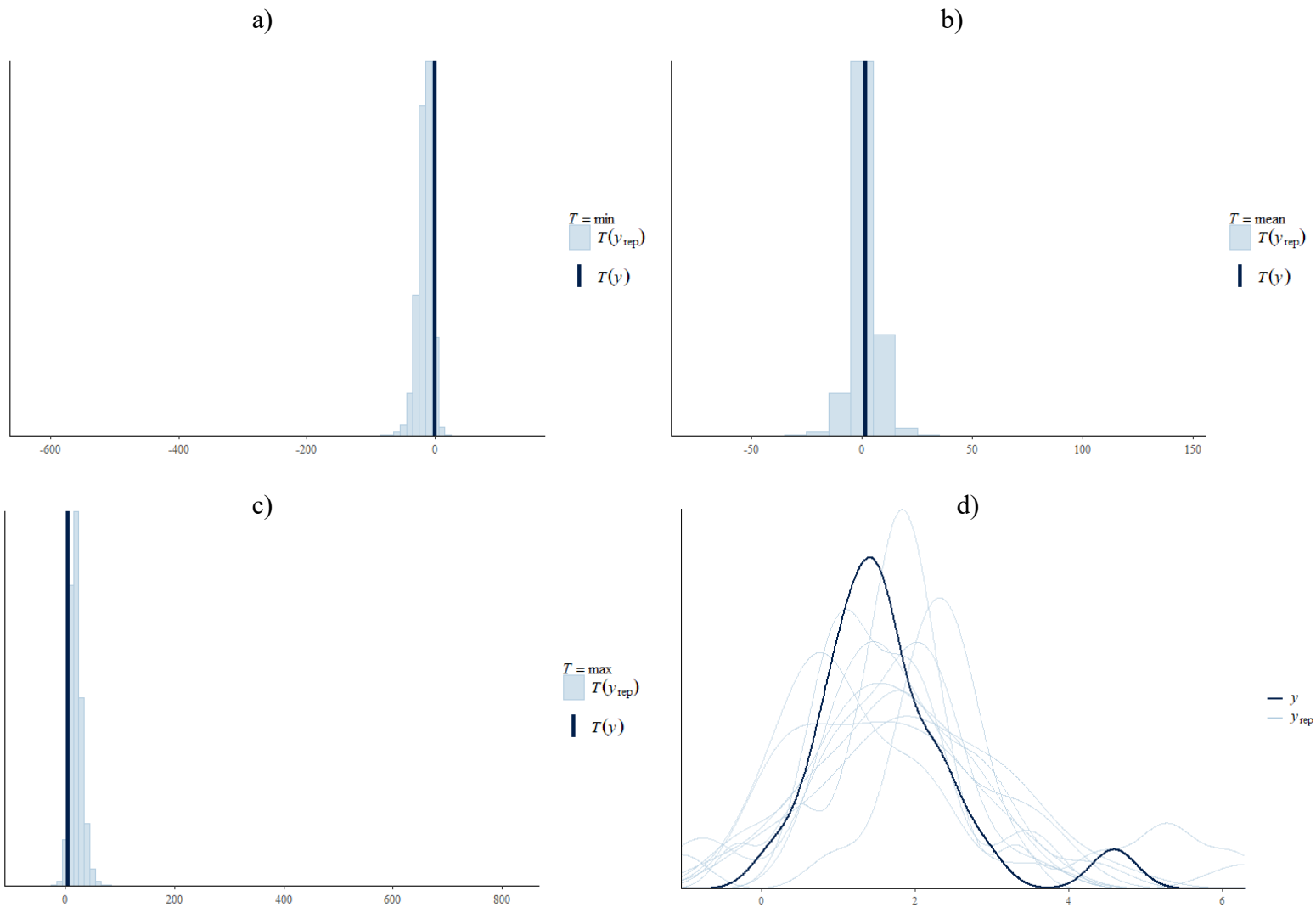



Figure 0.9: Prior predictive check of a model with N(0,1) for coiled bodies burden and N(0,10) for covariates, w-score ~ coiled bodies burden + imaging-death interval + gender (in dentate nucleus with full dataset). T(y) is the distribution of the data (dark blue line), T($y_{rep}$) is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

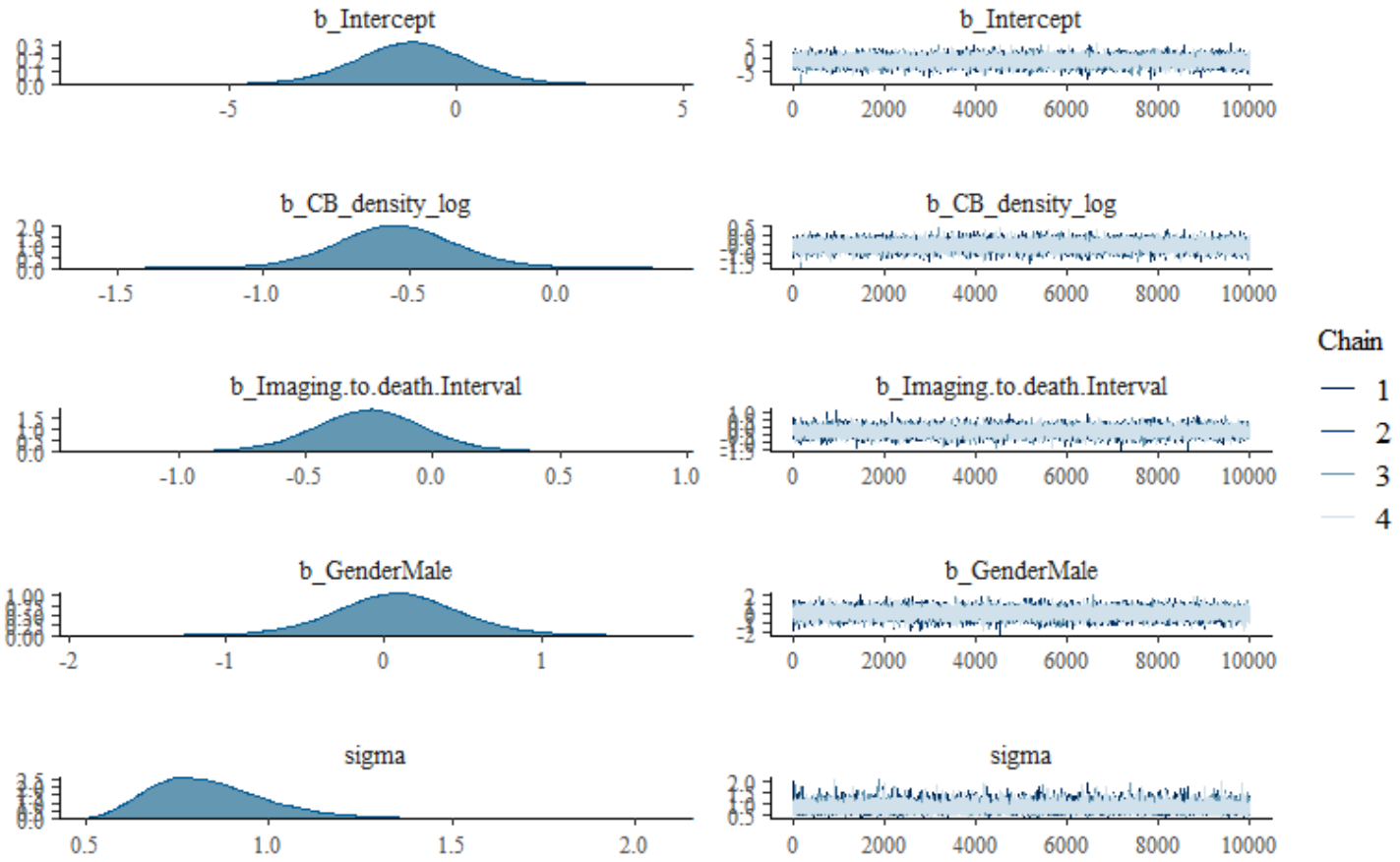



Figure 0.10: Plots of posterior distribution and trace plots of the regression coefficient in the model; *w-score ~ coiled body burden + imaging-death interval + gender* in dentate nucleus with full dataset. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.

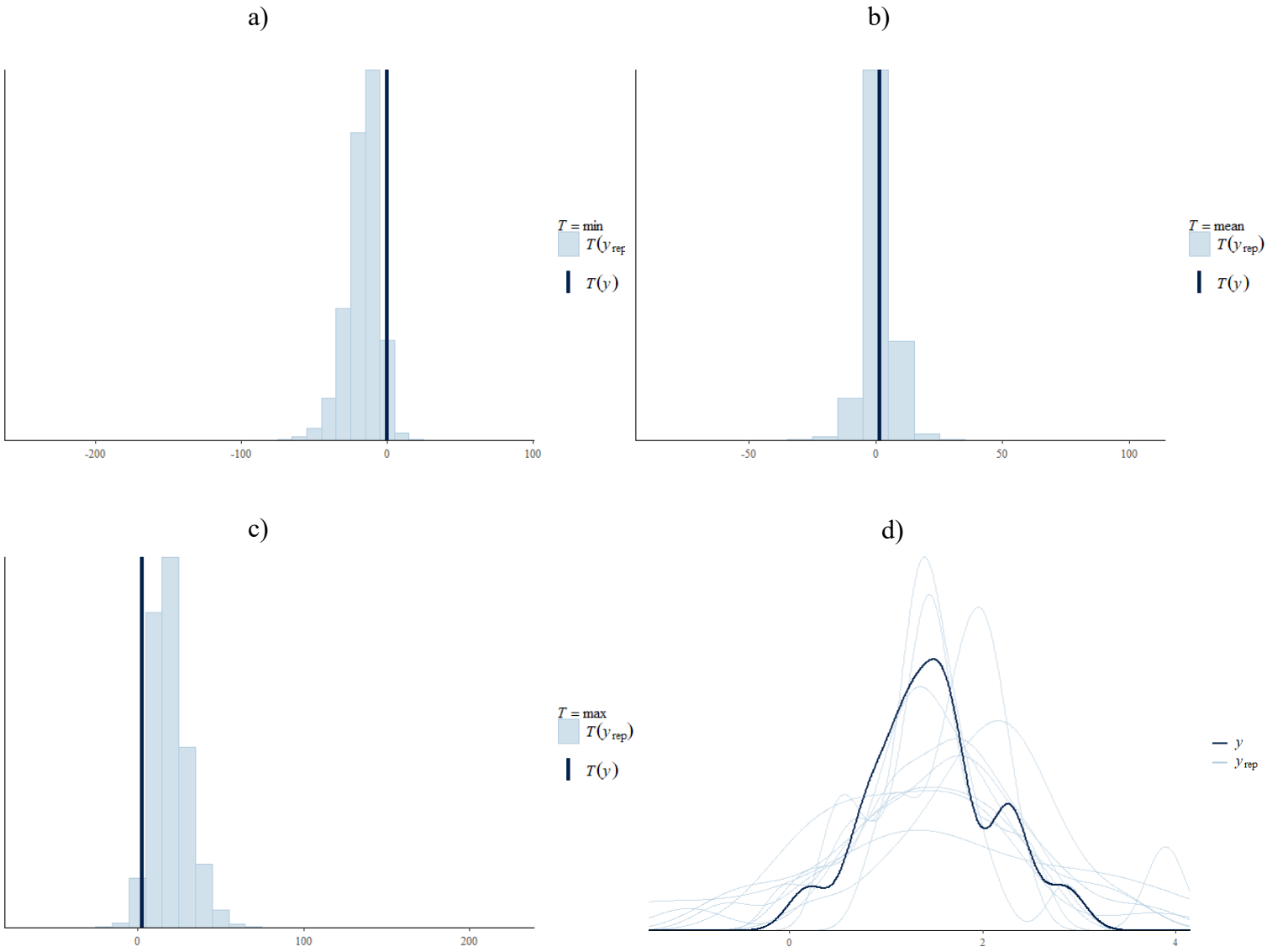



Figure 0.11: Prior predictive check of a model with N(0,1) for coiled bodies burden and N(0,10) for covariates, w-score ~ coiled bodies burden + imaging-death interval + gender (in dentate nucleus with the potential outlier removed). T(y) is the distribution of the data (dark blue line), $T(y_{rep})$ is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

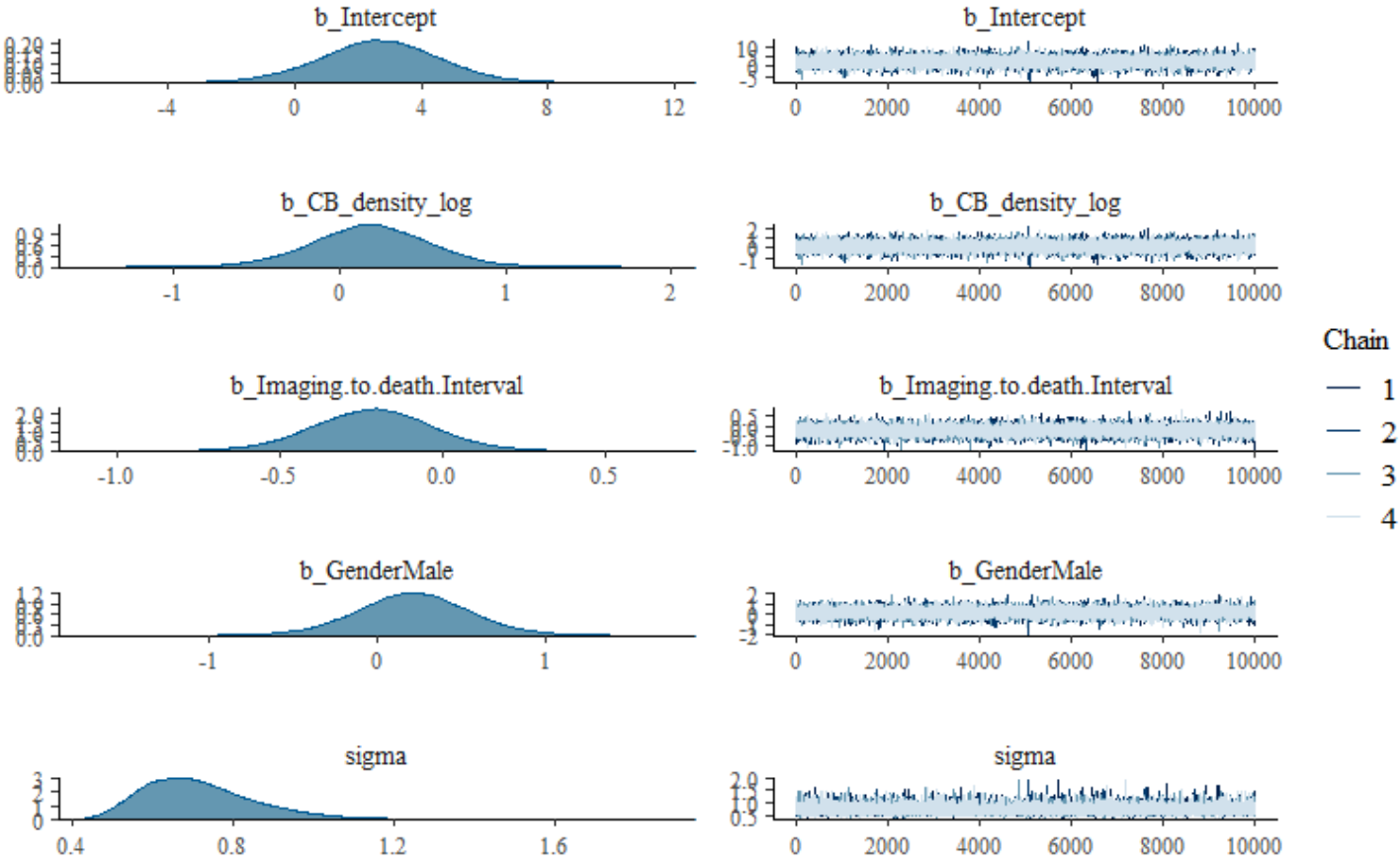



Figure 0.12: Plots of posterior distribution and trace plots of the regression coefficient in the model; *w-score ~ coiled body burden + imaging-death interval + gender* in dentate nucleus with the potential outlier removed. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.

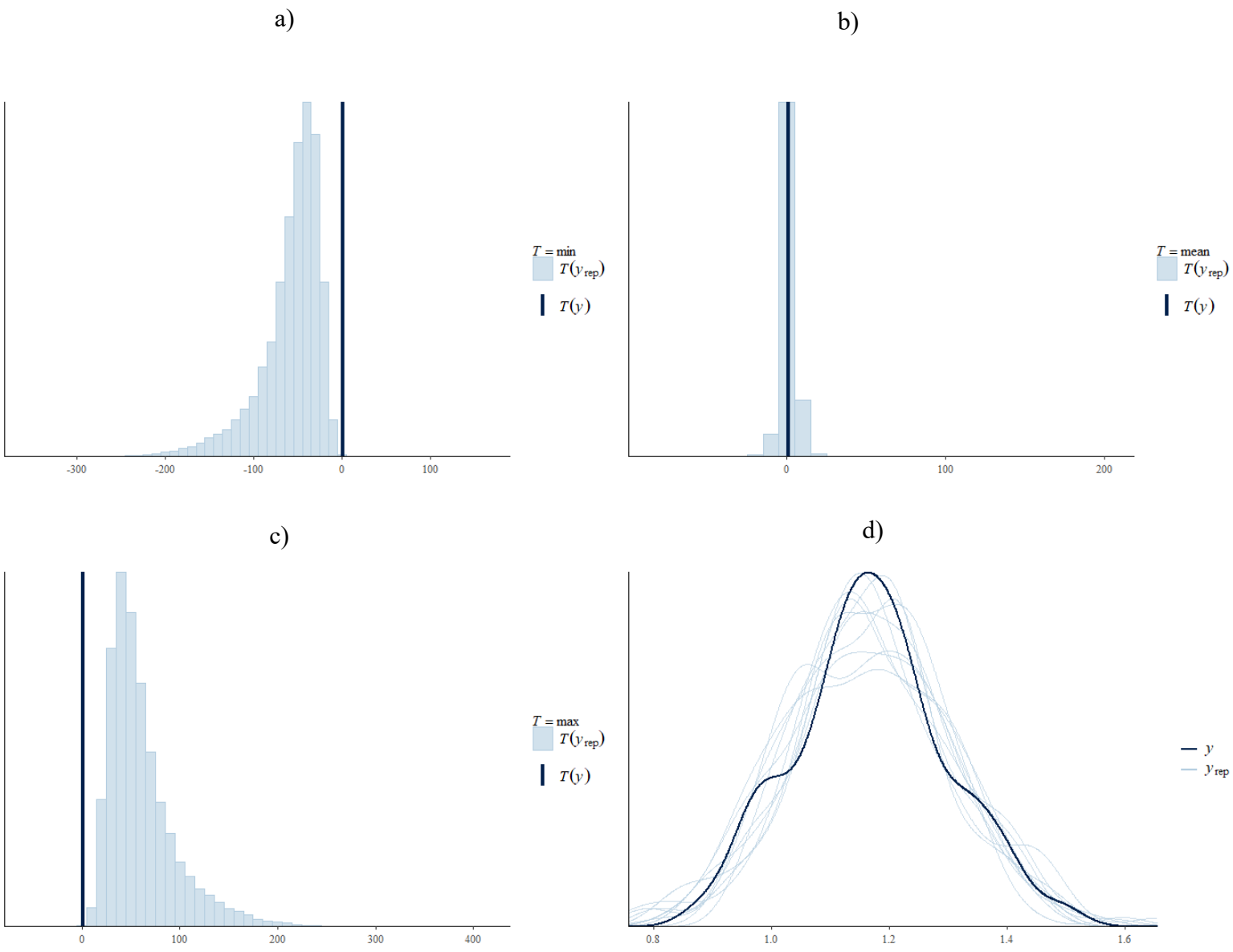



Figure 0.13: Prior predictive check of a model with N(0,10) as the prior distribution, weighted degree ~ tau density + imaging-death interval + gender + disease duration + echo type + (1|region name) + (1|Patient ID) across all regions. T(y) is the distribution of the data (dark blue line), T($y_{rep}$) is the data generated from the chosen prior (light blue). A chosen prior is appropriate when the dark blue line is more than minimum (a), within mean (b) and less than the maximum values (c) generated. Posterior predictive check (d) is shown where actual data (y) is plotted with the simulated data from the posterior distribution ($y_{rep}$). Actual data should be in-line with the simulated data.

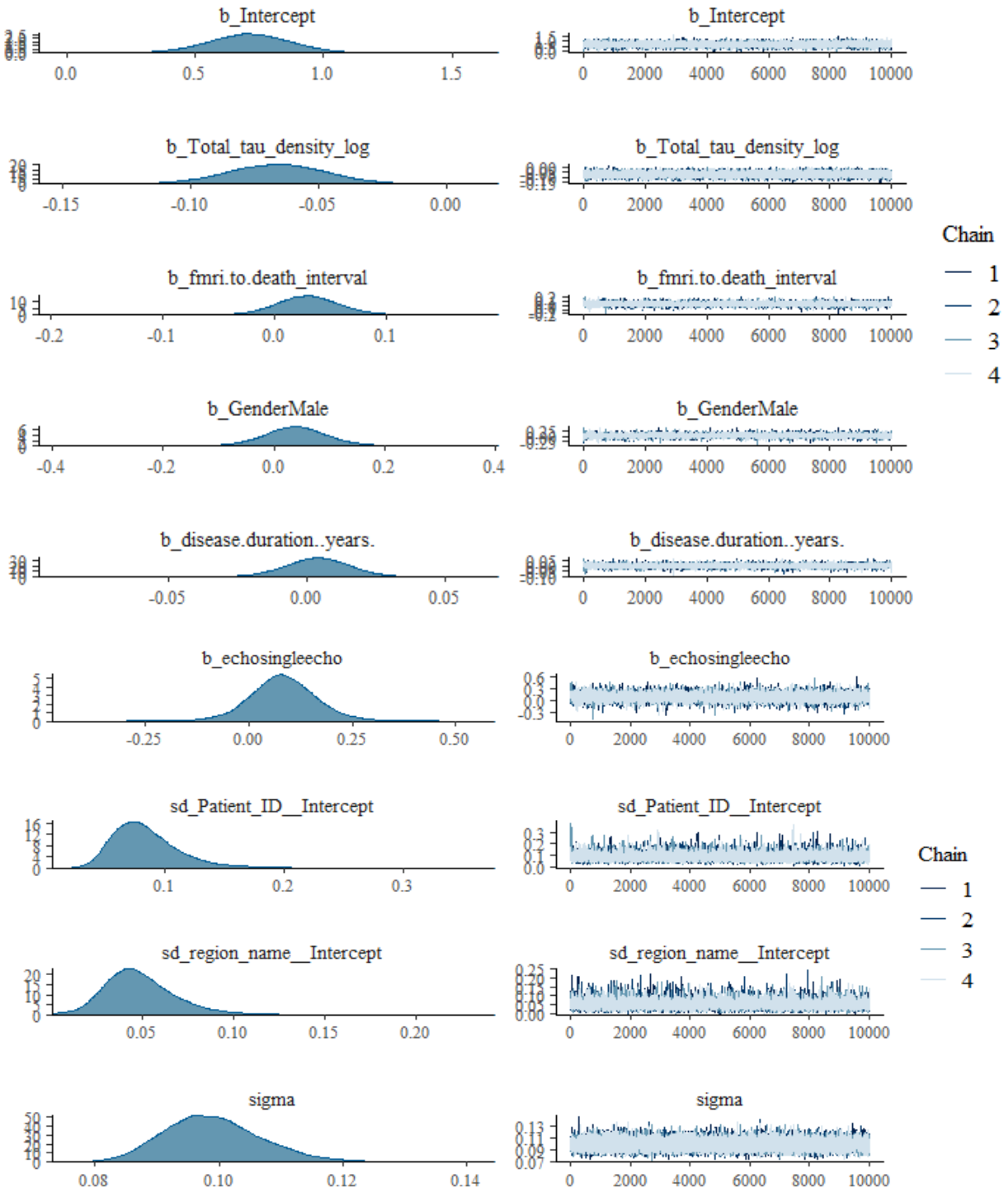



Figure 0.14: Plots of posterior distribution and trace plots of the regression coefficient in the model, weighted degree ~ tau density + imaging-death interval + gender + disease duration + echo type + (1|region name) + (1|Patient ID) across all regions. When the model is fit properly, trace plots should have no specific pattern observed, but should show random scatter around the mean.