Discovery of Plasma Biomarkers for Parkinson's Disease and Associated Cognitive Decline

Khosousi Fakhrabadi, Shervin

Awarding institution:
King's College London

Discovery of Plasma Biomarkers for Parkinson's Disease and Associated Cognitive Decline

Thesis submitted for the degree of
Doctor of Philosophy by

Shervin Khosousi

May 2023

Department of Old Age Psychiatry
Institute of Psychiatry, Psychology & Neuroscience
King's College London

Table of Contents

Table of Tables

Table of Figures

Abbreviations

List of abbreviations used throughout the thesis. Gene abbreviations not listed.

4PL	Four parameters logistic
4R-Tauopathy	4-repeat (4R)-Tauopathy
5PL	Five parameters logistic
ACC	Accuracy
ACN	Acetonitrile
AD	Alzheimer's Disease
AGC	Automatic gain control
ALS	Amyotrophic Lateral Sclerosis
ANOVA	Analysis of variance
APD	Atypical Parkinsonian Disorder
AUC	Area under curve
BBB	Blood Brain Barrier
BDI-II	Beck's Depression Inventory II
BH	Benjamini-Hochberg
CBD	Corticobasal Degeneration
CBS	Corticobasal Syndrome
CH50	50% haemolytic complement
CID	Collision-induced dissociation
CJD	Creutzfeldt-Jakob Disease
CNS	Central Nervous System
COMT	catechol-O-methyltransferase
CRP	C-reactive protein
CSF	Cerebrospinal fluid
CV	Coefficient of variation
DAT-SPECT	Dopamine transporter single-photon emission computed tomography
DLB	Dementia with Lewy Bodies
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FB	Factor B
FD	Factor D/Adipsin
FDR	False Discovery Rate
FH	Factor H
FI	Factor I
FTD	Frontotemporal Dementia
FTMS1	Orbitrap spectra
GWAS	Genome Wide Association Study
H&Y	Hoehn & Yahr
HADS	Hospital Anxiety and Depression Scale
HC	Healthy Control

HCD	Higher collision-induced dissociation
HPLC	Ultra-high-performance liquid chromatography
IAA	Iodoacetamide
IEF	Isoelectric focusing
IL-6	Interleukin-6
IPG	Immobilized pH gradient
ITMS2	Ion trap fragmentation spectra
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB	Lewy Body
LBD	Lewy Body Dementia
LCMS	Liquid Chromatography Mass Spectrometry
LEDD	Levodopa Equivalent Daily Dose
LOESS	Local polynomial regression
LogFC	Log2 fold change
LOOCV	Leave one out cross validation
MAC	Membrane Attack Complex
MADRS	Montgomery-Åsberg Disease Rating Scale
MAO-B	Monoamine Oxidase B
MBL	Mannose Binding Lectin
MCI	Mild Cognitive Impairment
MCP-1	Monocyte chemoattractant protein 1
MDS	Movement Disorder Society
MDS-UPDRS	Movement Disorder Society's Unified Parkinson's Disease Rating Scale
MFS	Mental Fatigue Scale
MoCA	Montreal Cognitive Assessment
MRI	magnetic resonance imaging
MRM	Multiple Reaction Monitoring
MS	Mass Spectrometry
MS	Multiple Sclerosis
MSA	Multiple System Atrophy
MW	Molecular Weight
NfL	Neurofilament light chain
NMS	Non-motor symptoms
NMSQ	Non-Motor Symptoms Questionnaire
NPH	Normal Pressure Hydrocephalus
OD	Optical density
PBMC	Peripheral Blood Mononuclear Cells
PCA	Principal component analysis
PD	Parkinson's Disease
PDCI	Parkinson's Disease with Cognitive Impairment
PDD	Parkinson's Disease Dementia
PDMCI	Parkinson's Disease with Mild Cognitive Impairment
PDND	Parkinson's Disease with No Dementia
PDQ-39	Parkinson's Disease Questionnaire 39
PPV	positive predictive value

PRM	Parallel Reaction Monitoring
PSP	Progressive Supranuclear Palsy
PSQI	Pittsburgh Sleep Quality Index
pTau	Phosphorylated Tau
QC	Quality Control
RBD	REM Sleep Behaviour Disorder
REM	Rapid Eye Movement
RFE	Repeated Feature Elimination
RGP	Resorufin-D-galactopyranoside
ROC	Receiver operating characteristic
RT-QuIC	Real-time quaking-induced conversion
SABC	HRP-Streptavidin Conjugate
SBG	Streptavidin-β-galactosidase
SCD	Subjective Cognitive Decline
SDS	Sodium dodecyl sulphate
SIMOA	Single Molecule Array
SNpc	Substantia nigra pars compacta
SPE	Solid phase extraction
SPS	Synchronous Precursor Selection
SRM	Selective Reaction Monitoring
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
SVM	Support Vector Machine
TBI	Traumatic Brain Injury
TBM	3,3’,5,5’-Tetramethylbenzidine
TCEP	tris(2-carboxyethyl)phosphine
TCS	transcranial sonography
TEA	Triethylamine
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
TLR	Toll Like Receptor
TMT	Tandem Mass Tag
TNF-α	Tumour Necrosis Factor α
t-SNE	t-distributed stochastic neighbour embedding
tTau	total Tau
YLK-40	Chitinase 3-like 1
α-syn	α-synuclein

Declaration of Thesis

I hereby declare the work within this thesis is my own, unless where explicitly referenced to work of others. The contents are original, and I have not submitted the thesis or any work therein for a degree at another university.

Covid Impact Statement

My PhD studies took place between June 2019 and August 2022. Covid restrictions of varying degrees have been in place between March 2020 and April 2022, including two longer national lockdowns. For the first nine months my PhD work was not impaired by Covid, and during that time I performed the optimising experiments for the mass spectrometry studies and learnt wet lab and data science skills. PhD students were not allowed into the lab between March and August 2020, which delayed the proteomic discovery experiment. Owing to increased safety regulations, additional approvals and paperwork needed to be in place before I was allowed back into the lab, which took another two months. Additionally, every time I was not allowed in the lab at King's College London, I returned to Sweden, where my co-supervisor Per Svenningsson has his lab. Returning to UK during Covid has also been problematic, owing to travel restrictions and quarantining rules. Furthermore, every time I was in contact with a Covid positive, or tested positive, I had to self-isolate and was not allowed into the lab. A second lockdown took place in the winter of 2021 with a surge in cases with the Delta variant, which again restricted access to the lab. Luckily, I had finished sample preparation for the mass spectrometric study by then. The mass spectrometry core facility was considered an essential facility that remained open throughout the lockdowns, and my samples were able to be processed.

Between the spring 2021 and spring 2022, there were different levels of restrictions, but access to the lab was possible, and work from home was largely encouraged.

However, the focus of my PhD had to change a bit because of the circumstances, and I have focused more on *in silico* analysis, and less on wet lab than originally planned.

Covid has also restricted in person supervision, courses, conferences, and networking. Online learning and communication over Zoom/Teams have been possible, but less effective than their in-person counter parts.

Acknowledgements

I would like to thank everyone who supported and supervised me during my PhD. Colleagues, friends, and family have all been indispensable these years.

I would like to thank my whole supervisory team for their guidance. I want to give a special thanks to Abdul Hye who has been exceptionally helpful these years, taking on the role as my day-to-day supervisor and teaching me all aspects of biomarker research. He has moreover been incredibly supportive in my PhD as a whole and has been present as a mentor and friend. I also want to give a particular thanks to Per Svenningsson who has been very supportive and made this PhD possible, through funding, samples, contacts, guidance, and supervision. Moreover, I want to thank Latha Velayudhan for acting as my primary supervisor and guiding me through my PhD.

I also want to thank the proteomic facility at Denmark Hill, both Steven Lynham and Xiaoping Yang have been helpful with both analysing our samples and answer any proteomic related questions I have had.

I am also grateful for having been surrounded by fantastic fellow researchers and PhD students these years, both in the Wohl and during my visits to Sweden.

This thesis would not have been possible without all of them.

Abstract

Parkinson's disease (PD) is a common neurodegenerative disorder diagnosed by the presence of bradykinesia, in combination with either rest tremor, rigidity, or both. Additionally, many non-motor symptoms (NMS) such as dementia, constipation, anosmia, depression, and sleep disorders are common. The clinical presentation of PD is heterogeneous, and increasing evidence suggests PD is rather a syndrome of parkinsonism with varying aetiologies. One of the most debilitating features of PD is the high prevalence of cognitive impairment, which has been linked to pathology spreading to limbic and cortical regions. A definite PD diagnosis can only be established post-mortem, where the loss of dopaminergic neurons and detection of protein aggregates containing α-synuclein called Lewy Bodies (LB) confirm the diagnosis. There is moreover evidence that the periphery is affected in PD. Clinically, many NMS often involve peripheral systems, including sensory and autonomic dysfunction, and systemic organelle and immune dysfunction are implicated in PD. Biomarkers are lacking in PD, both for early accurate diagnosis, and to differentiate PD endophenotypes, such as which patients will develop dementia. Previous plasma biomarker discovery studies in Alzheimer's disease (AD) have shown promising result using mass spectrometry.

The aim of this thesis was to attempt to develop a pipeline approach for biomarker discovery and verification, in order to find novel plasma biomarkers for PD with and without cognitive impairment. A large-scale untargeted mass spectrometry experiment was performed using plasma from PD patients with varying degrees of cognitive impairment and healthy control (HC). The data was analysed for diagnosis, disease severity, cognitive impairment, and cognitive decline, and a list of candidate biomarkers was generated. The most robust peptides were selected for a targeted mass spectrometric experiment to verify the candidate biomarkers. Neurofilament light chain (NfL), arguably the most well studied PD biomarker today, was quantified in plasma as a benchmark. Finally, a validation study was performed for complement factors using immunoassays, as they are the most implicated differentially expressed protein group in PD plasma.

2260 proteins were quantified in at least half the study participants. Over 50 plasma proteins were differentially expressed in PD, and the most implicated pathway was the complement and coagulation pathway. 17 proteins had an absolute Cohen's d effect size larger than 0.6. A single protein could differentiate PD from HC with a 76% accuracy, whereas a panel of 10 protein increased the diagnostic accuracy to 90%. Additionally, many candidate biomarkers were found for severity of cognitive impairment and longitudinal conversion to cognitive impairment and dementia. 20 proteins were

significant both for level of cognitive impairment and cognitive decline. The most robust and reproducible peptides were selected for a targeted proteomic verification study. In total 22 peptides from 12 proteins were quantified, and 6 proteins correlated well with the discovery study. APOC3, C9, and TGFBI were successfully verified for PD diagnosis, and SAA1 correlated with motor and cognitive severity. Additionally high plasma LUM and APOA4 were both associated with dementia conversion for PD patients. NfL levels were elevated in PD plasma, and were pronounced in the cognitively impaired patients, and heavily affected by older age. Comparatively, some of the novel PD biomarkers identified by mass spectrometry are arguably better candidate biomarkers for PD diagnosis and PD cognition. An attempt was made to validate complement system biomarkers, but no differences were found between PD and HC, however, a decrease in C1q and C3 was found in atypical parkinsonian disorders.

In conclusion, the discovery study generated many novel plasma biomarker candidates for both PD diagnosis and cognition. The pipeline was successful in verifying several plasma proteins discovered in the untargeted PD biomarker study. Many novel biomarkers from the discovery study, and some from the verification study, showed superior biomarker performance to NfL. However, a limited number of proteins had robust enough peptides to be verified with the targeted mass spectrometric method used, and future studies should attempt using other techniques to verify the remaining candidate biomarkers.

1. General Introduction

1.1 Parkinson's Disease

1.1.1 Clinical presentation and diagnosis

Parkinson's disease (PD) is the second most common neurodegenerative disorder after Alzheimer's disease (AD) [1]. PD is diagnosed clinically based on motor symptoms, commonly using the Movement Disorder Society's (MDS) guidelines [2] where the main criteria are bradykinesia, in combination with either rest tremor, rigidity, or both. Other criteria are often considered for the correct differential diagnosis, including response to dopaminergic medication, and absence of symptoms that would suggest Atypical Parkinsonian Disorders (APD). Symptoms that would exclude a PD diagnosis are for example cerebellar abnormalities, supranuclear gaze palsy, frontotemporal symptoms, treatment with dopamine antagonists. A definite pathological PD diagnosis can however only be established post-mortem, where the loss of dopaminergic neurons and detection of protein aggregates called Lewy Bodies (LB) confirm the diagnosis [3]. Clinical diagnosis has been shown to be sometimes be inaccurate, and a study from Mayo clinic found that only 77% of patients clinically diagnosed with PD (without dementia) had Lewy body pathology [4], whereas the majority of the remaining 23% comprised atypical parkinsonian disorders.

The motor symptoms are classically what characterise PD and are as mentioned still used today as the diagnostic criteria [2]. Bradykinesia, the mandatory motor symptom according to the MDS, is defined as slowness of movement with a decrease in speed or amplitude of movement and is primarily evaluated in the limbs of the patient. Rigidity is characterised by notable resistance or slowness of passive movements in major joints during physical examination. Resting tremor is defined as a low frequency (4-6Hz) tremor in a resting limb. Another common, although not diagnostic, motor symptom in PD is postural instability, which often occurs in later disease stages.

Despite the diagnosis being set based on motor symptoms, non-motor symptoms (NMS) are just as prevalent among PD patients and are responsible for much of the patients' disease burden [5]. The list of NMS associated with PD is long and could roughly be split into sensory dysfunction, neuropsychiatric symptoms, sleep disorders, and autonomic dysfunction [6]. Sensory dysfunctions include pain, visual and olfactory impairments, and sometimes appear before motor symptoms. Neuropsychiatric features are common in PD, and often co-exist with each other. These include depression, anxiety, fatigue, and cognitive impairment, and can appear at several different stages of the disease. Sleep disorders affect most PD patients in some form, and rapid eye movement (REM) sleep behaviour disorder (RBD) is one of the most accurate clinical predictors for developing PD.

Finally, autonomic dysfunction includes constipation, bladder dysfunction, and orthostatic hypertension, all of which suggest an involvement of the peripheral nervous system in PD [6].

1.1.2 Heterogeneity of Parkinson's Disease

The clinical presentation of PD is heterogeneous, and increasing evidence suggest PD is perhaps not just one disease, but rather a syndrome of parkinsonism with varying aetiologies [7]. There have been a few suggestions of how to categorise PD patients based on their clinical presentation [8]. One common approach is to classify patients by motor symptoms. PD patients are said to be either tremor dominant, postural instability and gait disorder dominant, or akinetic-rigid dominant [9]. This way of subtyping has however been criticised, as it assumes the main pathological difference between PD phenotypes lies the their motor presentation and does not account for NMS [10]. Moreover, both prospective and retrospective studies have shown that the motor subtype PD patients were classified as initially rarely remained the same a few years later. Additionally, unsupervised cluster analyses of PD patients based on their clinical symptoms do not tend to split the patients into these motor phenotypes [10].

Data driven cluster analyses have since tried to de novo subtype PD based on clinical presentation [8, 11]. Several of these studies have suggested NMS are stronger predictors of more severe disease outcome, such as faster progression and cognitive decline, whereas motor dominant phenotypes were associated with more benign outcomes. One data driven cluster study identified three new clusters that they defined as mainly motor/slow progression, diffuse/malignant, and one intermediate cluster. The mainly motor/slow progression group presented as the name suggests mainly with motor symptoms and had less rapid disease progression [11]. The diffuse malignant group on the other hand had a higher level of cognitive impairment, orthostatic hypotension and RBD at baseline. This has led to hypothesis on body-first and brain-first subtypes of PD [12] which refers to where the pathological changes with α-synuclein aggregation and propagation originates. The bottom-up type would theoretically start with pathology in the periphery that enters the central nervous system (CNS) for example via the vagal nerve, and affect pontine structures first followed by the basal ganglia. These would be the patients that first present with degeneration of cardiac sympathetic and gut parasympathetic nerves, reflected both by pathological evidence, and development of constipation, orthostatic hypotension and RBD before onset of motor symptoms [8]. The brain-first subtype would conversely be patients with initial limbic predominant pathology that present with early motor symptoms. Based on this, Berg et al. [8] propose a model of PD subtypes (Figure 1) based on prodromal clinical presentation, how the patients progress, and the relationship with known genetic causes for PD.

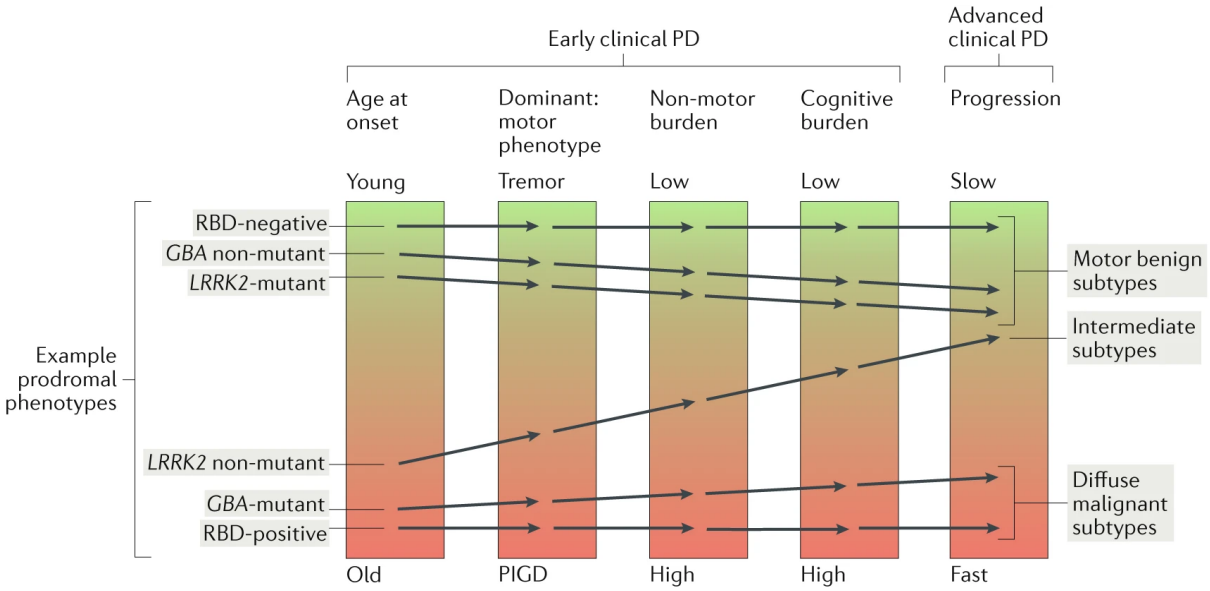

FIGURE 1 – OVERVIEW SUBTYPES OF PD (FIGURE ADAPTED FROM *PRODROMAL PARKINSON DISEASE SUBTYPES*, BERG ET AL., 2021 [8])

1.1.3 Aetiology end epidemiology

The incidence of PD diagnosis is higher in the elderly population, and older age is the largest demographic risk factor for developing PD. Approximately 1% of the population over 60 years, and 3% of those over 80 years of age are diagnosed with PD, and the number is increasing with a higher life expectancy [13, 14]. PD is however not uncommon in younger individuals, with 5-10% of patients receiving their diagnosis before the age of 50. Sex differences have been demonstrated in PD with approximately 1.5 times greater disease incidence in males [15], and presentation of symptom is slightly different between sexes. Men present more often with cognitive impairment and earlier motor symptoms than in women. Women present more often with tremor first, and develop dyskinesias, anxiety and depression more frequently [15]. Environmental risk factors have been difficult to pin down, but there is some evidence certain toxins, pesticides, infections, and head trauma possibly increase the risk of future PD. Documented protective factors on the other hand include smoking, coffee consumption, and anti-inflammatory drug use [16].

1.1.3.1 Genetics

Majority of PD cases are considered idiopathic, but at least 10-15% have been linked to risk genes [17], the most common being *SNCA* (Synuclein Alpha), *LRRK2* (Leucine Rich Repeat Kinase 2), *PRKN* (Parkin RBR E3 Ubiquitin Protein Ligase), *PINK1* (PTEN Induced Kinase 1), and *GBA* (Glucosylceramidase Beta). Although these are some of the most established risk genes for PD, Modern genome wide association

studies (GWAS) have identified mutations in over 20 genes associated with the disease and 90 independent risk-associated variants [18-20]. Mutations in *SNCA* [21], the gene coding for α-syn, was the first mutation reported to cause autosomal dominant PD [22]. *SNCA* mutations are associated with earlier disease onset, faster progression and higher prevalence of dementia and psychotic symptoms. *LRRK2* is strongly linked with PD, particularly the autosomal dominant Gly2019Ser mutation, and is clinically indistinguishable from idiopathic PD [23]. It is also the most common cause of familial PD globally. Interestingly *LRRK2* has also been associated with nigral neurodegeneration without concomitant Lewy body pathology. Mutations in *PRKN*, encoding parkin protein that is involved in the ubiquitin-proteasome pathway, have been linked with juvenile and early onset PD and slower disease progression [7]. It was the second PD gene to be identified, and spreads in an autosomal recessive manner. An associated gene, *PINK1*, encoding a mitochondrial serine/threonine protein kinase, binds parkin and is associated with mitochondrial dysfunction and mitophagy [24]. Finally, *GBA* gene variants pose an increased risk for PD and are present in 8.5% of patients, and linked with more rapid cognitive decline in PD. *GBA* carriers have a well characterised clinical phenotype, and commonly have an earlier disease onset, present with asymmetric resting tremor, and have severe motor impairment [22].

1.1.4 Treatment

The treatment of PD consists of replacing dopamine extrinsically, is mainly symptomatic, and aimed at improving motor function. Although new medications and means to improve dopamine function have come about, the treatment principle has not changed much over the past 50 years [25], as there is still a lack of disease modifying medication. A cornerstone treatment for many PD patients is levodopa that is converted into dopamine in the body. Levodopa tablets are normally combined with a dopamine decarboxylase inhibitor, such as benserazide, that does not cross the blood brain barrier and prevents levodopa to dopamine conversion outside the CNS [26]. Although initially effective, many patients develop motor fluctuations and dyskinesias (involuntary movements) over time. One can often mitigate these symptoms with Monoamine Oxidase B (MAO-B) or catechol-O-methyltransferase (COMT) inhibitors that slow down the breakdown of dopamine. An alternative treatment option is dopamine agonists, that mainly act on the D2/D3 system. Although it works well for some patients and reduce dyskinesias, there is an increased risk of psychiatric side effects, such as impulse control disorders, psychosis, and confusion [26]. Non-dopaminergic medication is mainly given to PD patients to alleviate NMS, such as serotonin/norepinephrine reuptake inhibitors for depressive symptoms, and acetylcholinesterase inhibitors for apathy and dementia [27]. There are currently no disease modifying treatments available for PD, but several are being investigated [28, 29].

Many current drug trials aim at reversing pathology associated with PD, for example through α-syn reducing therapies, restoration of mitochondrial dysfunction, immunomodulation, targeting genetic causes of PD, iron chelating medication, as well as many existing drugs that are being repurposed for PD.

1.1.5 Pathology

1.1.5.1 Central Nervous System

Numerous pathological mechanisms have been implicated in PD, including organelle dysfunctions, inflammation, protein aggregation and vesicle transport disruption. Most likely a combination of some of these mechanisms give rise to the PD phenotype [7]. A main pathological attribute in PD is the loss of dopaminergic neurons, particularly in substantia nigra pars compacta (SNpc), which disrupts dopaminergic signalling in the nigrostriatal pathway. The dopamine disruption affects both the direct and indirect pathways of the basal ganglia, and is believed to cause the motor symptoms, particularly bradykinesia [30]. Neuronal loss can also be found in other dopaminergic pathways [31], and even affect other neurotransmitter systems. Noradrenergic, serotonergic, and cholinergic pathways are also impaired in PD brains. The neurodegeneration is accompanied by intracellular aggregates called Lewy bodies (LB), which mainly consist of the protein α-synuclein (α-syn) which has been misfolded. PD pathological hallmarks of LB and neurodegeneration as outlined by the Braak stages [32] commonly start in the peripheral nerves and the medulla and propagate in a caudal-to-rostral direction to the pons, midbrain, limbic system, and eventually to cortical structures [33] (Figure 2).

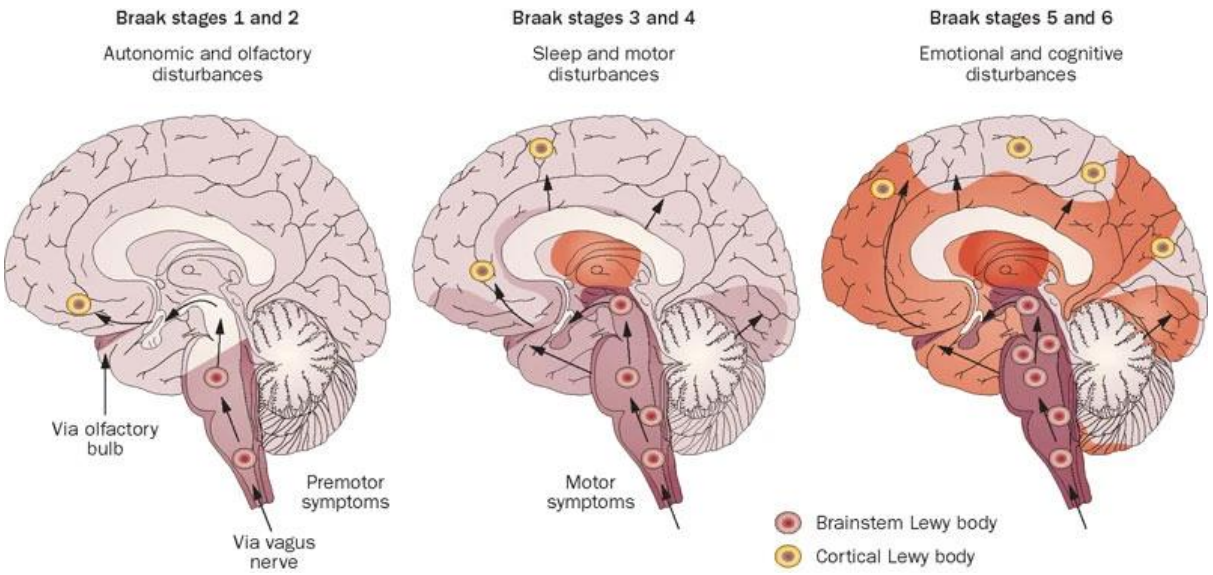

FIGURE 2 - BRAAK STAGING OF LEWY BODY PATHOLOGY IN PARKINSON'S DISEASE (FIGURE FROM *OLFACTORY DYSFUNCTION IN PARKINSON'S DISEASE*, DOTY, 2012 [33])

1.1.5.2 α-synuclein

Lewy bodies are seen in nearly all individuals with PD, and their main constituent is misfolded α-syn. α-syn is a 140 amino acid long protein physiologically highly expressed in neurons, but also in other tissues such as bone marrow, erythrocytes, kidney, and skin [34]. It is mainly localised in the presynaptic terminals, but when forming LB under pathological conditions it is also found in the cell soma [34]. Its function is not clear, and studies are trying to elucidate what makes the protein aggregate and spread in neurons according to the Braak stages in the first place. α-syn left long enough on its own in high concentrations will form fibrils, which propagate LB like inclusions in different *in vitro* and *in vivo* experiments [35, 36]. α-syn in the body exists in an equilibrium between soluble monomeric form and multimeric membrane bound state [37]. α-syn is constituently phosphorylated and dephosphorylated *in vivo*, which changes the kinetics of the protein, and phosphorylated α-syn appear to inhibit aggregation of α-syn [37]. Moreover, it has been seen that α-syn can spread between neurons via several mechanisms, including transmitted in vesicles, taken up from other dead neurons, as well as through axons. Given the *in vitro* propensity of α-syn to self-aggregate together with evidence of its transmissibility, some studies even suggest α-syn has prion like properties [38].

1.1.5.3 Periphery

There is ample evidence that the periphery (outside the CNS) is affected in PD. Clinically, many NMS often involve peripheral systems, including sensory and autonomic dysfunction [6]. Many of these symptoms have been attributed to PD pathology spreading through peripheral nerves, particularly through cranial nerves. α-syn pathology is seen in the skin [39], enteric nervous system, vagal nerve

[40], and the olfactory bulb [41] of PD patients. Additionally, more widespread systemic changes not limited to the nervous system are seen in PD, and include mitochondrial [42], lysosomal [43], and inflammatory dysfunction [44].

A central role of the immune system has been linked to PD in many studies [45]. Epidemiologically, individuals with autoimmune disorders have a 33% higher risk of developing PD, according to a large-scale Swedish study [46], and some diseases with a strong inflammatory component such as inflammatory bowel disease have increased risk of developing PD [45]. The blood brain barrier (BBB) has long been considered to render the CNS immune privileged, but this view has since been revised. The brain is conversely an immune-specialised organ that interacts with the peripheral immune system across the blood brain barrier directly or through the lymphatic and glymphatic systems [44]. There are several lines of evidence for increased microglial activity in PD brain, such as PET studies with higher binding of radioligands to activated microglia in PD [47] and evidence of increased number of activated microglia in post-mortem brain [48] . Some studies suggest misfolded α-syn activate microglia via the Toll-like receptors (TLR) TLR4 or TLR2 and cause aberrant inflammation that damage vulnerable cells such as dopaminergic neurons [44]. Moreover, T-cells have been shown to infiltrate PD brain, and express proinflammatory phenotype peripherally [49]. Some aspects of the immune defence, such as the complement system, are less studied in PD but have indirect evidence for their role. For example, several proteomic studies claim complement components in plasma are some of the most differentially expressed peripheral proteins in PD [50-53]. Moreover, several immunohistochemical studies have found deposit of complement factors in LB [54].

Mitochondrial dysfunction has also been strongly associated with PD pathology [42], although the exact mechanisms are still unknown. The hypothesis originally emerged when toxins inhibiting complex I in the mitochondrial electron transport chain were associated with dopaminergic loss and parkinsonism, and mitochondrial dysfunction was subsequently noticed in cells of PD patients. Later, several common genetic causes of PD have been linked with mitochondrial function. *PINK1* and *PRKN* are both necessary for normal mitophagy, and code for key proteins that bind the mitochondrial outer membrane and are involved in ubiquitination of depolarised mitochondria [42]. Other genetic causes of PD have also been connected to mitochondria. *LRRK2* is for example involved in transport of damaged mitochondria, and *LRRK2* mutations have been associated with hampered mitophagy [55]. Moreover, α-syn appears to disrupt normal mitochondrial function on its own though several mechanisms [56].

Autophagic breakdown of intracellular components via lysosomes is crucial for normal cellular function, and postmitotic cells such as the majority of CNS neurons are particularly sensitive to disruptions of the autophagic system [57]. α-syn is degraded though this system, and studies suggest

both alterations in α-syn protein as well as inhibition of the autophagy system led to accumulating α-syn. Furthermore, several genetic risk factors have been connected to lysosomal function, including mutations in the *GBA* gene that code for GCase [58]. Additionally, decreased GCase activity was also found in PD without *GBA* mutation. Several other genetic mutations for lysosomal proteins have also been linked with PD, and individuals with lysosomal storage disorders generally have a higher incidence of neurodegenerative disorders.

The gut is a peripheral organ system that has been associated with PD that has been extensively studied the past years. Up to a third of PD patients have gastrointestinal symptoms [59]. Constipation is the most common gastrointestinal symptom in PD patients, and interestingly often present many years before diagnosis. This together with evidence of α-syn propagation through neurons, and PD related pathology in the vagus nerve, has led to the hypothesis of PD originating in the periphery [40]. This theory has been backed up with studies showing individuals who had undergone vagotomy had significantly lower risk of developing PD [60]. This was further demonstrated in mice that were injected with α-syn fibrils in the gut that spread to the brain and caused PD like pathology and behaviour [36]. Moreover, vagotomy or α-syn knock-out prevented the spread of the pathology to the brain.

1.2 Parkinson's Disease Dementia

1.2.1 Clinical presentation

Cognitive decline is one of the most common NMS in PD and major cause of lower quality of life [61], and PD patients decline more frequently and at a higher rate than the normal population. The point prevalence of dementia in PD is reported to about 25-30 %, and up to 80% of PD patients are expected to decline cognitively at some point during their life [62]. Although cross-sectionally about a quarter of all PD patients have dementia at any one time [63], the rate of decline is highly variable on an individual level. While many patients progress rapidly and will be in need of full care, others do not seem to decline cognitively at all. Subjective cognitive decline (SCD) is the earliest phase and is a self-perceived decline in cognition with a normal cognitive test [64]. Mild cognitive impairment (MCI) is an early phase of PD dementia (PDD) with a mild decline in cognition. 10-20% of PD patients have MCI at diagnosis and often present with a faster cognitive decline, although this varies greatly. Multiple cognitive domains are affected in PDD, but two of the four executive abilities, attention, visuospatial abilities, and memory should be severely affected for a diagnosis. The PDD diagnosis is set when motor symptoms develop at least one year before onset of dementia [65], whereas an earlier onset of dementia would results in a Dementia with Lewy Bodies (DLB) diagnosis.

1.2.2 Pathology

Pathologically PDD is associated with Lewy bodies and a disrupted dopaminergic system not only in the basal ganglia, but also cortical and limbic areas of the brain. Moreover, concomitant amyloid pathology is often found, and the cholinergic system is affected [61]. PDD is considered a Lewy body dementia (LBD) and is diagnosed in a patient with an already established PD diagnosis with parkinsonism for at least a year prior to developing dementia [66]. The dopaminergic deficit is seen in the caudate nucleus in PDMCI cases, and in the limbic and neocortical regions in PDD. Similarly, noradrenergic, and cholinergic deficits are more widespread in the brains of PDD compared with PD with normal cognition.

1.2.3 Compared with other dementias (DLB)

It is often debated whether PDD and DLB are distinct disorders as their symptoms are overlapping and are only separated by the time point dementia appears in relation to parkinsonism [67]. On group level PDD cases have more advances motor symptoms and more affected dopaminergic pathways in the basal ganglia, whereas DLB patients have more cortical pathology and cognitive symptoms [67]. This makes sense as PDD patients develop motor symptoms earlier than DLB patients, whereas DLB

patients develop cognitive symptoms first. Moreover, classical AD fluid biomarkers are altered more in DLB compared with PDD, but this might be attributed to higher level of overlapping AD pathology [68, 69]. It is still debated whether cortical and limbic AD like pathology or Lewy pathology contributes more to cognitive symptoms seen in DLB. Recently there has been some mechanistic evidence suggesting the α-syn strain in DLB and PD/PDD are different and lead to different aggregates [70] and could explain the differences in the disorders.

1.3 Biomarkers

1.3.1 Role of biomarkers in PD

Accurate, specific, and reproducible biomarkers are valuable for all diseases, and are currently lacking for PD. Early diagnosis of PD is crucial for implementation of disease modifying drugs to salvage the progressively degenerating neurons before patients start to develop symptoms [71]. Biomarkers are also necessary for correct differential diagnosis, as many APD patients are incorrectly diagnosed initially [4], and there are few early predictors of which patients will develop PDD [61]. Moreover, prognostic biomarkers are essential to differentiate disease phenotypes [8], to use as endpoints in drug trials, as well as gaining insight in the heterogeneous disease pathophysiology. Ideally biomarkers should also reflect specific pathological processes that can be targeted with novel treatments.

PD biomarkers can roughly be subdivided into four main domains: clinical, imaging, genetic and biochemical biomarkers [71]. Clinical symptoms are the most convenient to assess and pose a useful initial approach to stratify patients. Besides using symptoms for diagnosis and endophenotypic classification, they can also serve a predictive purpose. Loss of smell, constipation, and sleep disorders (particularly RBD) are all associated with increased risk of developing PD [6]. Although many of these prodromal PD NMS are prevalent, they are either too unspecific or are not sensitive enough, hence cannot be used on their own as predictive biomarkers [72]. RBD has a high rate of PD conversion (>80%) [73], but only a minority of PD patients (around 25%) have RBD at time of diagnosis. Moreover, many individuals with RBD also develop DLB, and in some cases even MSA, which makes it non-specific. Hyposmia, which is a common prodromal symptom in PD was shown in one study to have a positive predictive value (PPV) of 12.5% to convert to PD in an elderly population [74]. Although this PPV is many times higher than that for the average population, it also means 87.5% of individuals with only hyposmia did not develop PD at all in the study. However, when paired with imaging, the PPV increased to 67%, which suggests that a panel or stage wise screening would potentially be the most useful. Many studies are being conducted on imaging biomarkers, both structurally with magnetic resonance imaging (MRI) and transcranial sonography (TCS), as well as functionally with radiotracers, mainly targeting the dopaminergic system. Dopamine transporter single-photon emission computed tomography (DAT-SPECT) for example shows reduced striatal binding in PD [75], and many MRI studies look at morphological changes as a proxy for neurodegeneration [76]. Some promising advances have been made in the MRI field where iron sensitive, neuromelanin sensitive, diffusion sensitive, and resting state functional magnetic imaging measures can capture changes in the nigrostriatal system of the PD brain [77]. Although most PD cases are idiopathic, several attempts have been made to identify genetic markers [20, 78]. These include both familial and genome wide association studies identifying mutations in *GBA*, *LRRK2*, *SNCA* among other genes. The genetic mutations found are primarily seen

as PD associated risk factors rather than biomarkers on their own. Finally, plenty of research has been conducted on biofluid markers in PD, where most of the studies have been targeted and hypothesis driven. Most have been conducted on either cerebrospinal fluid (CSF) or plasma [79], although other samples types (urine, peripheral blood mononuclear cells (PBMC), tissue biopsies, saliva etc.) have also been tested.

1.3.2 Biofluid biomarkers

Biofluids, or biological fluids, refer to a wide range of fluids produced by the body, that are either secreted, excreted, or surround and perfuse cells and organs [80]. Examples of the latter are blood, CSF, and lymphatic fluid. All of these comprise extracellular spaces of the organs, regulate the biological environment, and transport waste and nutrients to and from cells [81]. This means biofluids often reflect ongoing healthy and pathological processes in the body. The circulatory system approximates all cells of the body, as cellular respiration is dependent on exchange of oxygen and carbon dioxide as well as energy nutrients [82]. Fluids escaping the vasculature of the circulatory system, end up in the extracellular space and is absorbed by the lymphatic system, which eventually re-enters large veins of the circulatory system [83]. This means that blood is an excellent source for detecting ongoing pathology, particularly as it is easily and routinely taken in clinics.

1.3.3 Biofluid markers in other neurodegenerative diseases

1.3.3.1 Alzheimer's disease

Major progress has been made in biofluid biomarker discovery and validation in neurodegenerative disorders, particularly in the past few years when ultra-sensitive technologies have enabled accurate quantification of CNS derived biomarkers in blood. The AD field has experienced some of the greatest advancements, with Neurofilament light chain (NfL) [84], phosphorylated tau (p-Tau) [85], and beta amyloid (Aβ) [86] now successfully reflecting AD in both CSF and plasma. Aβ in CSF is already established in clinics. Modern mass spectrometers as well as immunoassays are able to detect the similar changes in plasma [87], and commonly the ratio Aβ42/40 is decreased in AD and correlates with amyloid pathology seen in PET. Tau is another already established clinical CSF biomarker, both total Tau (t-Tau) and phosphorylated Tau (p-Tau) are increased in AD CSF [88]. p-Tau has recently proven to be a powerful plasma marker for AD. More than 40 phosphorylation sites have been identified for Tau, and some of these species (pTau181, pTau217, or pTau231) are significantly increased in AD [87]. pTau181 is the most extensively studied phosphorylated tau isotype, as it appears to discriminate AD from both healthy individuals as well as other dementias and other

tauopathies with high accuracy [89]. Moreover, pTau181 appears to be elevated early, and reflect AD brain pathology. More recently pTau217 gained attention as it appears to have an even higher accuracy than pTau181 [85]. Neurofilament light chain (NfL) is another biomarker that is elevated in both plasma and CSF, but it appears to reflect neurodegeneration in general rather than reflecting AD specific pathology [84]. Other altered promising biomarkers are synaptic proteins. Synaptic loss in AD is associated with cognitive decline and seen as a downstream effect of pathological AD mechanisms. The most studied one is neurogranin which is increased in AD CSF and correlated with cognitive decline and brain atrophy and seems to predict with future cognitive decline in individuals with MCI [90].

1.3.3.2 non-Alzheimer's disease

Plasma p-Tau species have been found to be moderately elevated in non-AD dementias but much less elevated than in AD patients and appear to reflect the amyloid pathology. A recent study comparing PD, DLB, AD, and control plasma pTau181 and pTau231 found a slight increase in PD patients, followed by DLB and the highest levels seen in AD, which did not overlap with the other disorders [68]. Moreover, a strong association was seen between plasma pTau181 and pTau231 levels with both CSF Aβ42 levels and cognitive scores. Apart from a moderate decrease in CSF Aβ42 for Lewy body dementias, core AD CSF biomarkers (Aβ42, t-Tau, p-Tau) appear to be like cognitively healthy controls for non-AD dementias [91]. NfL is a subunit of neurofilaments, a structural protein of neurons, particularly highly expressed in axons. NfL has gained popularity in many neurodegenerative disorders, owing to its strong correlation with axonal degeneration or injury [92]. The CSF levels are about 40 times higher than in plasma, and plasma levels can only be quantified accurately using ultra-sensitive methods such as the Simoa. The elevated levels of NfL have been studied in several disorders including HIV-associated dementia, Amyotrophic Lateral Sclerosis (ALS), Creutzfeldt-Jakob disease (CJD), Multiple Sclerosis (MS), APD, Traumatic Brain Injury (TBI), Frontotemporal Dementia (FTD), and Normal Pressure Hydrocephalus (NPH) in descending order. A more moderate elevation, however significant, is seen in AD and Lewy body dementias. Only a mild and not always significant increase can be seen in PD, but it seems to be associated with PD progression [93]. However, NfL is possibly useful to differentiate PD from APDs.

1.3.4 CSF biomarkers in PD

1.3.4.1 α-synuclein

Much of the CSF biomarker research in PD has been hypothesis driven, and many studies have therefore focused on measuring α-syn in CSF. α-syn is quantifiable using immunoassays, with decent inter assay correlation [94]. Multiple meta-analyses have assessed the use of total α-syn in CSF as a

PD biomarker, with similar conclusions that total α-syn is slightly lower in PD CSF compared with healthy controls [95-98]. However, the results between studies have been varied, with some studies showing insignificant changes, and high inter study variation in α-syn concentrations. Hence, although there is a concordance that α-syn is decreased in PD CSF on group level, it might not serve as a good biomarker for an individual patient. There appears to be a large individual variation in levels of CSF α-syn levels, and a pooled diagnostic specificity of around 50% [95]. Specific α-syn variants associated with PD and Lewy bodies have also been attempted as potential biomarkers [94]. Oligomeric α-syn, an intermediate pre-fibril form, has been found at higher concentrations in PD CSF. However, it is a transient α-syn form that rapidly regresses to monomers, or aggregate to form fibrils, and its diagnostic performance has been unsatisfactory to be used as a biomarker [99]. Around 90% of Lewy Body α-syn is phosphorylated [100], and phosphorylated α-syn has been measured to be elevated in PD CSF. The diagnostic accuracy of phosphorylated α-syn is generally moderate, similar to that of the oligomeric species [99].

Recently there have been additional advancements in using CSF α-syn as a PD biomarker. Real-time quaking-induced conversion (RT-QuIC), an *in vitro* seeding aggregation assay, has generated promising results with around 90% sensitivity and 100% biomarker specificity in some studies [101-103]. In the RT-QuIC assay, CSF samples are added to a buffer with recombinant monomeric α-syn, and samples containing aggregated or seeding forms of α-syn, such as fibrils, will generate more aggregates over time [101]. Besides having a good diagnostic performance against PD, it appears to be able to differentiate between different synucleinopathies such a DLB, PD and MSA [70, 104]. Studies have even reproduced findings in CSF using skin and mucosal tissue with good results [35].

1.3.4.2 Other targeted PD biomarker studies

A few other biomarkers related to known PD pathology have been measured with mixed results. Lysosomal dysfunction is implicated in PD, and mutations in the *GBA* gene are the most common genetic risk factors for developing PD, resulting in impaired lysosomal degradation glucoceramide. Glucocerebrosidase activity has been found to be lower in PD CSF, and a number of lysosomal markers appear to be associated with cognitive decline in PD [105]. Several lines of evidence have suggested the immune system plays a key role in development of PD [44]. Many studies have measured various inflammatory markers in the CSF, out of which Monocyte chemoattractant protein-1 (MCP-1) and Chitinase 3-like 1 (YLK-40) have come up as the most promising candidates, although results have been varying.

1.3.4.3 Classic AD markers in PD

Validated CSF biomarkers for other neurodegenerative diseases such as AD have been studied in PD. Lower levels of Aβ42 have been observed in PD in some studies, analogous to AD, however the reduction is more pronounced in PDD and DLB and appears to reflect amyloid pathology in patients [105]. Tau species (t-Tau and p-Tau) have also been investigated in PD, but with varying results [94]. NfL has been extensively studied in PD, as it has shown to be a good general neurodegenerative biomarker that is increased in many disorders. NfL in CSF appears to be elevated in APD and be useful as a differential biomarker to PD [104], but unchanged in PD versus controls. Some recent longitudinal studies however have observed higher NfL at baseline correlating with faster decline in motor symptoms and cognitive function [106, 107]. Moreover higher NfL levels were associated with greater overall PD symptom severity [93].

1.3.5 Blood/plasma in PD

1.3.5.1 Target PD blood biomarkers

To date, no blood biomarkers are established for PD, although numerous attempts have been made. Among targeted studies, α-syn has been the most extensively studied. A meta-analysis [108] found a general increase in total plasma α-syn, although there was a great heterogeneity in the cohorts and assays used. The effect size was also variable, where some studies found no difference, and others a 100-fold increase in PD [109]. Biomarkers that reflect other pathological PD mechanisms have also been studied. Inflammation has been implicated in PD with increased microgliosis in the CNS, and aberrant T-cell as well as monocyte activity peripherally [44]. This has resulted in many studies measuring plasma cytokines in PD. Generally, an increase is seen in many pro-inflammatory cytokines (such as Interleukin-6 (IL-6), Tumour Necrosis Factor α (TNF-α), and C-reactive protein (CRP)) [110], however these are not very specific, nor show a great discriminatory power to potentially be used as biomarkers. Another studied biomarker candidate is plasma urate, where large epidemiological studies have suggested higher urate levels have a protective effect against developing PD [111, 112]. Randomised controlled trials with urate elevating drugs have however not shown any disease modifying effects on PD progression [113].

1.3.5.2 Use of mass spectrometry to find biomarkers

A few studies have attempted discovering biofluid biomarkers for PD in an untargeted manner. A recent systematic review of 12 PD blood proteomic studies identified 23 candidate biomarkers [114]. Almost all these studies were based on 2D gel electrophoresis separation followed by mass spectrometry. Several of the biomarkers were consistent across studies, with Apolipoprotein A-I and

haptoglobin emerging as the most reproduced findings. Moreover, most differentiated proteins across studies were either apolipoproteins or complement proteins [114]. Despite this, there is limited literature on validation studies on candidate markers, nor studies that explore their function in relation to PD. There were however some technical limitations in the studies included in the review. Nearly for all studies the plasma or serum samples were depleted of the most abundant proteins (usually albumin and immunoglobulins), separated on a 2D gel, extracted, and digested, before analysing on a mass spectrometer. This is a labour-intensive protocol, and not sensitive to low abundant proteins that might not be visible on the gel. Hence, on average, less than 100 individuals were included per study, and half of them included less than 30 PD patients. Moreover, most proteins detected were highly abundant in blood, and the most sensitive studies managed to identify a few hundred proteins in the analysis. Given that a large portion of the proteins in the higher ug/ml - mg/ml concentration range are complement factors and apolipoproteins, it is not surprising many of the candidate markers belong to these groups. Furthermore, it is extremely difficult to detect CNS derived proteins with the techniques used, many of which are in the pg/ml range in plasma. NfL for example is about seven orders of magnitude less abundant in plasma compared with complement factor C3 [115, 116].

In recent years methods and equipment for mass spectrometric discovery studies have improved detection and quantitation of the plasma proteome. In a study from 2019 Ashton et al. [117] used plasma from 284 individuals from two independent cohorts to discover novel biomarkers to predict amyloid burden in preclinical AD cases. They TMT labelled the samples to minimise inter-run variation, and pre-fractionated the samples with an orthogonal method to the HPLC. They detected a total of 2356 proteins, some present even in the pg/ml range. Furthermore, their top hits were neuronal proteins including NfL, Amyloid precursor protein (APP) and Neurogranin 2 (NGN2). This study showed that it indeed is possible to discover plasma biomarkers for neurodegenerative disorders using mass spectrometry, but it has not yet been done in PD.

1.3.6 PD Dementia markers

Just as for diagnosing PD, biomarker studies have been performed trying to predict which PD patients develop cognitive decline. Clinical, genetic, imaging, and biofluid markers have been studied. Older age at diagnosis is one of the greatest predictors of rapid cognitive decline in PD patients [64]. Clinically non-tremor dominant variants of PD are more likely to decline cognitively, compared with tremor dominant forms [118]. A few other predictors, such as male sex, hallucinations, early severe anosmia, and RBD are all associated with cognitive decline in PD [119]. Voxel based MRI studies have demonstrated volume loss and increased rate in atrophy in mainly cortical and limbic/paralimbic areas

in PDD and PD-MCI [120]. This is in line with histological findings in the same regions. Some PET studies have furthermore found hypometabolism in certain cortical regions in PD is associated with future development of PDD [119, 121].

There are a few biomarkers which have altered levels in PDD compared with PD with normal cognition. Most of the well-studied biofluid markers in PDD have been derived from Alzheimer's disease (AD), with the most consistent finding being decreased amyloid beta 42 (Aβ42) in cerebrospinal fluid (CSF). A systematic review by Lin et al. [119] found that nearly all included studies had found a decrease of Ab42 in PDD CSF. the results on pTau and tTau were conflicting, where some studies showed moderate elevation, and others normal levels. However, other potential biomarkers have appeared in both CSF and plasma, such as increased proinflammatory markers as well as the free radical scavenger Uric acid [122]. The issue with all the PDD candidates so far is that although they are significantly changed in PDD, they have a low discriminatory power.

The role of alpha-synuclein (α-syn) in development of PDD has been studied. A meta-analysis on CSF levels of α-syn found comparable levels in DLB and PD, but were lower in both compared to AD [123]. They further found a correlation between level of cognitive impairment in DLB and α-syn levels, but this has not been demonstrated in PD/PDD. Recently a few groups have found correlations between NfL and cognitive decline in PD. One study found correlations between NfL levels and more rapid progression in both cognitive impairment and motor symptoms [107]. Survival analysis further showed high levels of NfL predicted conversion of PD to PD-MCI and PDD.

Few large-scale shotgun discovery studies have been conducted on PDD. One relatively large study from 2019 performed a discovery and validation study in plasma from 458 individuals using an aptamer method (SomaScan) [124]. They quantified 1129 proteins in plasma from individuals with PD and healthy controls in a discovery cohort, and successfully validated four out of the top ten most robust biomarker candidates in a validation cohort. Bone sialoprotein, osteomodulin, aminocyclase-1, and growth hormone receptor were all successfully validates, and growth hormone receptor was further associated with future cognitive decline. One high-throughput suspension bead array study quantified 216 proteins in PD CSF [125]. After verification of initial candidate proteins, kininogen1 was found to be associated cognitive performance in PD.

1.3.7 Strategies on finding neurodegenerative biomarkers

As seen from the studies described above, there have been many approaches for discovering new biomarkers in neurodegenerative diseases, some have been hypothesis driven and others untargeted. The hypothesis driven markers have mostly been based on known pathology, such as Aβ42 from

amyloid plaques, Tau from tau tangles, and NfL from axonal damage. Most of these proteins were originally only detectable in CSF, but after narrowing down candidates and protein isoforms from CSF studies, highly sensitive assays have been developed for blood where with similar results to CSF. Plasma has about 100 times the amount of protein compared with CSF [126], and generally lower concentrations for CNS proteins (around 50-fold lower for NfL and GFAP in plasma versus CSF). Hence ultra-sensitive methods such as the SIMOA are needed to accurately quantify these low abundant proteins in the complex plasma protein matrix.

Targeted approaches similar to AD have been attempted for PD, but with limited success. Proteins that generally correlate with neuronal and axonal death, like NfL, are barely elevated in PD. This is likely because there is not a rapid large-scale axonal degeneration in PD as is seen in other disorders. Much of the degeneration is limited to the dopaminergic neurons that comprise less than 1% of the brain's neurons. However, the degeneration is more widespread in more advanced disease where limbic and cortical areas are affected, and indeed NfL level do correlate with disease severity and cognitive decline [127].

The main histopathological hallmark of PD is Lewy bodies containing misfolded α-syn. Unfortunately, the findings for monomeric α-syn in PD CSF have not been as successful as CSF levels of tau and Aβ in AD. Studying other α-syn isoforms, such as phosphorylated α-syn or oligomeric α-syn, has yielded more promising results, although levels between PD and controls still overlap too much for clinical application. Moreover, translating it to a plasma assay is problematic, as α-syn is highly expressed in erythrocytes. α-syn levels are higher in plasma than CSF, which further complicates quantification of pathological α-syn derived from dying neurons in the CNS.

1.3.7.1 Using a discovery approach

An alternative way of finding new biomarkers is with an untargeted approach. Recent advancements in proteomic techniques have made it possible to detect low abundant proteins in biofluids even in untargeted studies. There are generally two approaches to untargeted proteomics, biased and unbiased methods. Biased approaches include proteomic techniques where a large number of proteins are quantified, but the proteins are preselected. These techniques include proximity extension assays (eg Olink [128]), bead assays (eg Luminex [129]), aptamer assays (eg SomaScan [130]), and well as some targeted mass spectrometric approaches such as parallel reaction monitoring (PRM) [131]. An advantage of these techniques includes being able to accurately quantify many proteins in one sample with a high sensitivity. As multiplexing technologies have improved, they pose a good alternative to classic mass spectrometric methods for discovering biomarkers. The Olink platform offers a 3000-plex in 2022, with assay sensitivities rivalling the most highly sensitive uniplex immunoassays available at the time. As stated, a disadvantage is that these methods are not

completely unbiased, and important proteins might be missed, however multiplexing methods that screen thousands of proteins still cover a significant portion of the quantifiable proteome.

Truly unbiased studies include gel separation and mass spectrometric approaches, but these also have their advantages and disadvantages. When properly pre-processed, thousands of proteins can be detected and quantified in human biological samples using mass spectrometry [117]. However, these studies are often very costly, and require highly specialised equipment. In highly sensitive mass spectrometry studies, the samples are processed much more than they would be for an immunoassay, which complicates back translation to original protein level [132]. For example, studies on plasma may begin with abundant protein depletion, followed by sample digestion, mass tag labelling, fractionation, and desalting. The peptides detected and quantified are then probability matched with a proteome database. This workflow introduces many potential confounders, and requires rigorous standardisation of the protocol, and ideally internal and external controls. Therefore, traditionally, many proteomic findings have been difficult to validate using orthogonal methods.

1.3.7.2 Building a pipeline to narrow down candidates

Using a shotgun proteomic approach for biomarker discovery is in many ways the ideal initial step for an unbiased biomarker screening. With sufficient sample number and pre-processing, one can detect and quantify thousands of proteins in an unbiased manner [117]. However, given the methodical caveats, it warrants subsequent verification and validation of candidate markers. Not all studies incorporate a verification step of proteomic biomarker candidates, but move straight to validating in an external cohort or using a uniplex immunoassay [133]. This has led to a high attrition rate of biomarker candidates [134], and the biomarker discovery pipeline in neurodegenerative disorders could be improved. One technique for verification of candidate biomarkers is to run the discovery experiment in two separate cohorts and see which candidate biomarkers overlap [135]. This would account for batch and cohort effects and reduce false discovery rates. An alternative strategy would be to run a large well phenotyped cohort in one single experiment, and cross validate biomarkers *in silico*. Modern statistical and bioinformatic methods allows for testing robustness of biomarkers with regularisation methods [124], biomarker performance through assessing accuracy and predictive value, and the effect of covariates (both biological and experimental). Moreover, machine learning methods can build prediction models on training data to test on another subset of data. These methods should be utilised, as a more stringent selection of consistent, specific, and significant biomarkers would be selected for experimental verification and validation [136].

The abovementioned verification techniques do however not eliminate the technical discrepancy between a mass spectrometry discovery study and immunoassay validation. Since high level of sample manipulation is necessary to reach the resolution for quantifying low abundant proteins, particularly

in plasma, using additional verification steps to narrow down candidate markers would be helpful. One common recommended approach is using selective reaction monitoring (SRM) or multiple reaction monitoring (MRM) [137], a mass spectrometric method that detects and quantifies individual peptides from a sample based on preselected mass spectra. As SRM is a targeted method, and the mass spectrometric setting can be optimised to the peptides quantified, the sample does not have to be processed as heavily to be quantified. Moreover, as the peptides monitored are preselected, and the entire proteome does not have to be probability matched with a database, the accuracy is generally much higher than an untargeted approach. Recently, parallel reaction monitoring (PRM)[138] has allowed for detection of multiple proteins at once in a targeted manner with more complete product ion coverage than with SRM. This is potentially a powerful multiplexing method, that also quantifies the same peptides that are detected in a discovery study, but with higher quantification accuracy [133]. SRM and PRM further allows for addition of standards, to gain absolute quantities of the peptides. A few studies in other disease areas have utilised this approach, and successfully validated candidate markers using mass spectrometry in biofluids [139-142].

1.3.7.3 Accounting for cohort heterogeneity

As mentioned, PD is a heterogeneous disorder with various clinical presentations. Several PD endophenotypes have been proposed and linked to genotypes and most likely have differing underlying pathologies [8, 11]. It is however not yet established which the main endophenotypes of PD are, which complicates the construction of an ideal study cohort. For now, one could therefore attempt to build a cohort with a wide range of clinical and demographic phenotypes for a discovery study, which would then enable endophenotypic analysis and correction for covariates. Certain demographic variables are better studied than others, and appear to affect biomarker levels. Motor symptom severity in PD is often used in studies as an indirect measure for disease severity [143]. After all, the diagnosis and start of the disease is often set based on motor symptoms. Although there is evidence that PD pathology often occurs years before onset of motor symptoms [144] and many patients mainly progress in their NMS, motor symptoms still pose a tangible and well-studied clinical measure and should be included in a PD biomarker study.

Dementia is one of the most well studied NMS in PD, and several biomarkers have been correlated with cognitive decline in PD [145]. Currently, several dementia biomarkers from AD are tested for PD dementia with varying results [146].

Demographically, age and sex are almost always included as covariates in biomarker studies, and for a good reason. Age is the major risk factor for morbidity in general, and the plasma proteome changes

with age [147]. The incidence of PD and PD dementia increases with age, and matching or correcting for age is imperative for any biomarker study.

Sex is another covariate almost always accounted for in biomarker studies. Many diseases are more prevalent in either males or females, or present differently between sexes. The reason for this is likely due to a combination of differences in developmental physiology, immunology, metabolism, behaviour, and body structure to name a few [148]. Some diseases are clearly linked to sex-specific organ development, such as the higher incidence of breast and ovarian cancers in females, and more prostate cancer in males [149]. Other, less obvious differences in morbidity include higher prevalence of autoimmune diseases in females [150], and aortic aneurysms in males [151]. Among psychiatric disorders depression is more often diagnosed in women, and schizophrenia in males [152]. For neurodegenerative diseases AD is more prevalent in females and PD in males [153]. Biomarkers in AD, particularly pTau, appear to be associated with sex specific genetic risk, brain pathology, and rate of cognitive decline [154, 155]. Female DLB patients seem to present with more hallucinations, less sleep disorders, and have more mixed AD and LBD pathology compared with male DLB patients [156]. Within the PD population men present more often with cognitive impairment and earlier motor symptoms than in women, whereas women present more often with tremor first, and develop dyskinesias, anxiety and depression more frequently [15]. Biological differences are also seen between male and female PD patients. A transcriptomic meta-analysis of PD brain tissue found major sex differences in several key PD mechanisms such as inflammation, mitochondrial dysfunction, and oxidative stress [157]. Some studies even suggest a link between time of menopause in women and risk of developing PD [158]. Moreover, several biomarkers studies have revealed sex differences in PD [159]. The suggested protective effect of high urate levels seems to only be consistent for males. A meta-analysis on homocysteine found elevated levels in blood to be associated with increased risk of developing both PD and AD, with higher levels in males, and phenotypic differences between male and female PD [160]. Moreover several studies have reported on sex-specific microRNA differences in PD patients [161]. This highlights the importance of correcting for sex, or analysing males and females separately, when performing proteomic studies.

1.4 Summary and aims

Discovery of PD biomarkers has been difficult, as most hypothesis driven studies have generally yielded insufficient diagnostic and prognostic biomarker performances, and proteomic studies have either lacked sensitivity or been difficult to validate [94]. Biomarker research in AD has in the past few years shown that most proteins in the CSF also can be accurately quantified in plasma [162]. Technical advancements in proteomics research have rendered mass spectrometry a promising method for discovery and validation of neurodegenerative plasma biomarkers [133]. Ideally a biomarker needs to be specific, reproducible, have a good diagnostic performance, implementable in clinics, and preferably reflect ongoing pathology.

PD is a heterogeneous neurodegenerative disease that presents with several motor and non-motor symptoms [7]. As outlined, the clinical heterogeneity is coupled with a wide array of biochemical, cellular, and genetic processes implied in PD pathophysiology. It is not unlikely that different biological processes and risk factors gives rise to the different disease phenotypes. It is hence imperative to have a well characterised disease cohort when studying biomarkers. Some biomarkers, such as proteins related to α-syn or dopaminergic neurons, could in theory reflect general PD processes and be used as diagnostic or progression markers. Indeed, the recently developed α-syn RT-QuIC is a biomarker that belongs to this category [101]. Other biomarkers might on the other hand reflect processes seen in PD subtypes, such as mitochondrial or lysosome dysfunction as seen in some genetic forms of PD, or gut dysfunction which is more common in cases where the disease is believed to start in the periphery. This latter category of biomarkers could potentially aid in detecting early disease processes before nigral neurodegeneration, and aid in development of precision medicine.

One strategy for finding new neurodegenerative biomarkers reflecting CNS disease processes is to perform a discovery study in CSF and subsequently validate candidates in plasma using highly sensitive targeted methods [137]. The advantages of this strategy are that proteomic quantification of low abundant proteins is easier is CSF owing to its less complex matrix, and CNS derived proteins would be present in much higher concentrations. A disadvantage with this strategy would be that one might miss important peripheral proteins. Additionally, CSF is generally less available than plasma, and discovery studies run the risk of being underpowered. Having sufficient sample size is particularly important when using sensitive shotgun proteomic methods in order to avoid overfitting and false positive and negative results. Given this issue with cohort size, and the many pre-diagnostic peripheral and systemic changes implied in PD including gut dysfunction, anosmia, immune system aberrations, mitochondrial dysfunction, lysosomal dysfunction etc., it might even be preferential to use plasma rather than CSF to discover new PD biomarkers.

One of the most debilitating PD endophenotypes is the development of cognitive impairment and dementia. PD patients develop dementia 2.5-6 times more frequently compared to the general population [64]. About half the patients have developed dementia 10 years after diagnosis, but the rate of cognitive decline varies significantly. Several clinical and genetic risk factors have been identified for PD cognitive decline, and more limbic and cortical pathology is seen in these patients.

The aims of this project were to discover and verify plasma protein biomarkers for PD diagnosis and PD cognitive function with mass spectrometry. The Swedish Biopark cohort was used [163], from which individuals with a clinical PD diagnosis selected with varying levels of cognitive function, both cross sectionally and longitudinally. Additionally, age and sex matched healthy individuals were included. A pipeline strategy for biomarker discovery and verification was developed as outlined in Figure 3. A shotgun proteomic experiment was done to start with, which was optimised for detection of quantifiable proteins in plasma, based on a protocol from a previous study by the group [117] with a few modifications. Subsequently the data was processed, and the most robust and promising candidates selected for verification, based on how well they reflected PD diagnosis and PD cognitive decline. Finally, a targeted proteomic method was set up where the most robust peptides from the candidate biomarker proteins were quantified in digested plasma, to verify findings from the shotgun study. In parallel with this, a group of proteins (complement factors) were quantified in PD plasma that have previously appeared as candidate biomarkers in blood proteomic studies, in an attempt to verify the findings with immunoassays.

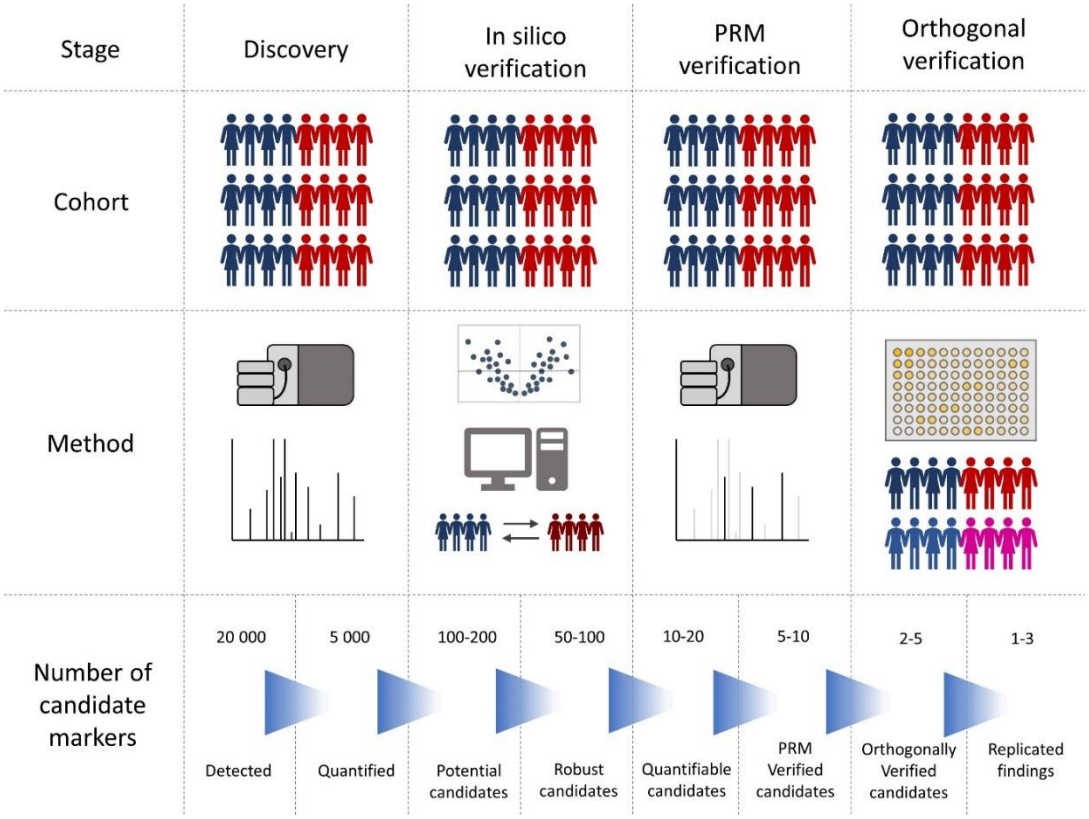

FIGURE 3 – SCHEMATIC OVERVIEW OF SUGGESTED WORKFLOW PROTEOMIC DISCOVERY AND VERIFICATION

Suggested workflow to narrow down proteomic candidate biomarkers. An untargeted large scale proteomic experiment is conducted on a large cohort. This generates a large number of protein IDs which need to be narrowed down in silico. A methodological verification step such as with parallel reaction monitoring (PRM) would confirm quantities of some of the peptides. The verifiable markers would then be validated using orthogonal methods, and ultimately tested in other cohorts.

1.5 Objectives

To find novel plasma protein biomarkers for Parkinson's disease

A) Discover new plasma biomarkers for PD and PD cognition

Conduct an untargeted experiment for novel plasma biomarkers discovery in PD with and without cognitive impairment using mass spectrometry

1. Optimise a protocol for analysing plasma on the mass spectrometer
    a. Investigate different protein fractionation methods
    b. Investigate different protein depletion methods
2. Discover biomarkers for Parkinson's disease (PD)
    a. Identify diagnostic biomarker candidates for PD
    b. Identify (motor) severity biomarkers for PD
3. Discover biomarkers associated with PD cognition
    a. Identify biomarkers associated with cognitive severity in PD
    b. Identify biomarkers associated with cognitive decline in PD
4. Select most promising biomarker candidates from the untargeted study to verify

B) Verify plasma biomarker candidates for PD and PD cognition

Conduct a targeted experiment for verification of plasma biomarker candidates

1. Use targeted mass spectrometry for more accurate relative quantification of candidate markers
    a. Compare results in untargeted with targeted study
    b. Assess biomarker performance of verified proteins/peptides
2. Evaluate the usefulness of the Discovery – Verification workflow

C) Quantify a current gold standard biomarker for neurodegeneration (NfL)

1. Assess the usefulness of NfL as a PD diagnostic and PD cognitive biomarker in plasma
2. Compare NfL with novel findings from the mass spectrometry study

D) Validate a category of candidate markers from earlier PD biomarker studies

1. Identify a category of biomarker candidates from previous PD plasma proteomic studies from the literature
2. Explore this groups of proteins (complement factors) as potential PD plasma biomarkers
    a. Use an orthogonal method (immunoassays) to validate a few complement factors as PD biomarkers
    b. Explore the general change in complement proteins in PD blood
        i. Explore functional changes in PD serum complement system
        ii. Explore relation of PD plasma complement factors with clinical PD symptoms

2. Methods

2.1 Cohort

2.1.1 Overview Biopark

The study participants used in all studies in the thesis were from the Swedish Biopark cohort managed by Stockholm Regional Council and led by Prof. Per Svenningsson. Biopark is primarily a Parkinson's disease cohort and has included over 700 PD patients by 2022, and patients are primarily seen in clinics within Stockholm. The cohort furthermore recruits healthy control individuals, as well as atypical parkinsonian disorders to a lesser extent. All participants in the study have given verbal and written consent. The study was approved by the Swedish Ethical Review Authority (Dnr 2016/19-31/12 and 2019-04967) and conducted under the Declaration of Helsinki.

Subjects are assessed at baseline, and at follow up time points 2, 5, 8, 11 years after baseline. Baseline is set at the time of inclusion in the cohort. For many patients that was around the time of diagnosis, whereas for patients that were diagnosed elsewhere before included in Biopark, baseline was set at date of inclusion.

At each time point the patients were assessed clinically by a neurologist as well as receiving self-assessment questionnaires.

PD patients were assessed with the Movement Disorder Society's Unified Parkinson's Disease Rating Scale[164] (MDS-UPDRS) parts I-IV, Hoehn & Yahr, and Montreal Cognitive Assessment[165] scales (MoCA). Where Furthermore, patients in both cohorts were self-assessed using a battery of scales including Beck's Depression Inventory II[166] (BDI-II), Montgomery-Åsberg Disease Rating Scale [167] (MADRS), Hospital Anxiety and Depression Scale[168] (HADS), Mental Fatigue Scale [169] (MFS), Pittsburgh Sleep Quality Index [170] (PSQI), Parkinson's Disease Questionnaire 39 [171] (PDQ-39), and Non-Motor Symptoms Questionnaire[172] (NMSQ). Disease duration was calculated from the date of diagnosis, and the Levodopa Equivalent Daily Dose (LEDD) was calculated according to a formula[173].

Short description on the clinical scales assessed that were included in the thesis:

- MDS-UPDRS is the most frequently used scale for assessment of PD and seen as the gold standard. It comprises of four parts, Part I: Non-Motor Aspects of Experiences of Daily Living; Part II: Motor Aspects of Experiences of Daily Living; Part III: Motor Examination; Part IV: Motor Complications. Some of the patients were assessed using an older version of the UPDRS scale,

and part III was used after converting the score using a conversion formula [174]. Generally, all patients were assessed ON medication, unless otherwise specified.

- The Hoehn & Yahr disease staging scale is part of MDS-UPDRS Part III, and a common way of assessing symptom severity. Patients can be classified on a scale from 0-5, 0 being asymptomatic, and 5 either wheelchair bound or bedridden.
- MoCA is a commonly used formula for cognitive decline assessing Visuospatial/Executive, Naming, Memory, Attention, Language, Abstraction, Delayed Recall and Orientation. It is often preferred in PD as it is more sensitive for mild cognitive impairment compared to for example the mini mental state exam.
- MADRS is a well-established scale for assessment of depression. It measures severity of depression related symptoms such as sadness, sleep, concentration difficulties, lassitude, pessimistic and suicidal thoughts. Although MADRS is good for assessing major depressive disorder as it incorporates many different domains, it might not be ideal for Parkinson's disease depression, as it overlaps with some PD symptoms (sleep disorders) that are not necessarily depression related.
- BDI-II is a psychometric test well suited for screening of depression but also measuring depression severity. It focuses more on the psychological symptoms than MADRS and may be better suited for PD depression.
- HADS is a self-rating scale for both anxiety and depression. It is a relatively short scale that is often used as a screening tool and focuses on mood and emotions. It does not include all criteria for depression and focuses on non-physical symptoms.
- NMSQ is a relatively new scale, used to cover many different NMS in one single questionnaire. It consists of 30 yes or no questions on sleep, cognition, perception, memory, sexual function, urological and gastrointestinal symptoms etc.
- PDQ39 is a PD specific scale that covers eight domains, including difficulties in daily living, as well as the impact of PD on specific functioning and wellbeing. It is well studies and validated, but not all encompassing as it does not cover sleep difficulties and sexual dysfunction.
- PSQI is a frequently used scale for assessment of sleep quality the past month. It is a reliable and valid scale, and includes sleep quality, sleep latency, sleep duration, and habitual sleep efficiency.

- MFS is a scale assessing mental fatigue and fatiguability. Beside mental exhaustion, it covers emotional, sensory, and cognitive symptoms. The scale has a high reliability for different patient categories.

2.1.2 Sample collection and preparation

Blood samples were collected with venepuncture within Region Stockholm by healthcare personnel. Plasma was prepared from blood collected in Ethylenediaminetetraacetic acid (EDTA) tubes and centrifuged through a density gradient medium (LymphoPrep™, cat#1114547) at 800 x g for 20min to separate peripheral blood mononuclear cells (PBMC) and plasma according to manufacturer's instructions. In short, EDTA blood was diluted 1:1 with phosphate buffer saline (PBS). 20ml diluted blood was layered on top of 15ml Lymphoprep in a 50ml tube and centrifuged at 800 x g for 20 minutes at room temperature. The top cell free plasma layer was collected.

Serum was prepared from whole blood collected that was centrifuged at 2000 x g for 15 min at 4°C. Both plasma and serum were aliquoted and stored at -80°C until further use.

2.1.3 Cohort for mass spectrometric studies

2.1.3.1 Cohort size

It is difficult to determine an optimal cohort size for a discovery study, as many factors affect how many study participants to include. Large variation in the quantified variables, large number of variables, smaller expected effect sizes, and higher cohort heterogeneity all warrant a larger cohort size. On the other hand, factors such as cost, labour, time, and sample availability often restrict the number of samples analysed. This makes traditional power calculations quite difficult, and the best way to determine sample size is often to look at similar studies. Ashton et al. published in 2019 a similar plasma proteomic study in AD, where they managed to identify verifiable candidate markers with a cohort size of 144 in the discovery study [117]. Our discovery study in PD has a similar study design, only with a few differences to the AD study. The mass spectrometer is more sensitive, and a higher number of protein identifications are expected, hence a larger cohort size is needed to minimise risk of false positives. Secondly, the cohort used is well phenotyped, which comes with advantages and disadvantages. An advantage is that it allows for exploration and discovery of PD endophenotype biomarkers, such as dementia specific markers. It also allows for better matched group comparisons and facilitates accounting for covariates. The disadvantage is risk of over-exploration and overfitting of data to disease endophenotypes. Taking into account the abovementioned factors, and considering sample availability, 198 study individuals (130 PD and 68

healthy controls (HC)) were included in the study. The controls were age and sex matched with the PD patients as closely as possible. Demographics are summarised in Table 1.

2.1.3.2 PD cohort construction

PD patients were selected from the Swedish Biopark cohort [163] based on a few criteria.

Definite inclusion criteria included:

- Clinical diagnosis of Parkinson's Disease by a movement disorder specialist
- MoCA score as close as possible to sample collection date, maximum 1 year apart

Selection criteria based on sample availability:

- Availability of multiple MoCA scores, at least 1 year apart from the first MoCA score
- As broad range of cross sectional MoCA scores as possible
- As broad range of rates of cognitive declines as possible
- Sex and age match cognitively impaired and unimpaired patients where possible
- Most complete clinical assessment available
- Most complete self-assessment available
- Sample date as close to date of diagnosis as possible

Patients were generally not selected if

- Severe comorbidity present that would significantly affect cognition/PD symptoms to a great degree
- Patient history/data that would suggest cognitive impairment prior to PD diagnosis
- DLB/dementia diagnosis at time of PD diagnosis or within a year after PD diagnosis

Healthy controls were generally selected if

- Closely matched age and sex of PD cohort

Healthy controls were generally not selected if

- Clinical data or patient history suggested dementia or cognitive impairment
- Severe somatic disease burden was present
- Neurological disease was present

TABLE 1 – DEMOGRAPHICS FOR MASS SPECTROMETRY STUDY

Demographics of Parkinson's disease patients and healthy controls. Clinical scales and other parameters include Hoehn & Yahr, Unified Parkinson's Disease Rating Scale (UPDRS) III, Levodopa Equivalent Daily Dose (LEDD), Montreal Cognitive Assessment (MoCA), Beck's Depression Inventory (BDI) II, Pittsburgh Sleep Quality Index (PSQI), Montgomery Åsberg Depression Rating Scale (MADRS), Mental Fatigue Scale (MFS), Parkinson's Disease Questionnaire (PDQ) 39, Hospital Anxiety and Depression Scale (HADS), and Non-Motor Symptom Questionnaire (NMSQ). Age presented as mean (standard deviation), all other data as median (range).

	Parkinson's disease	*Healthy Controls*
Age	68.3 (9.5)	65.5 (9.0)
Sex (Female : Male)	58 : 72	33 : 55
Disease duration	3.4 (0-22)	-
Hoehn & Yahr	2 (1-4)	-
UPDRS III	25.5 (3-80)	-
LEDD	500 (0-2235)	-
MoCA	25 (9-30)	-
BDI II	10 (0-43)	-
PSQI	7 (2-18)	-
MADRS	7 (0-36)	-
MFS	10.5 (0-32)	-
PDQ 39	25.8 (0-76.6)	-
HADS Anxiety	5 (0-19)	-
HADS Depression	3 (0-17)	-
NMSQ	8 (0-20)	-

2.1.3.3 Assessment of cognition and cognitive decline in PD

The PD group was primarily built to allow assessment of cognitive impairment and cognitive decline in PD. Several models of cognitive decline were considered for the available data. Ultimately, the main aim was to identify which patients declined cognitively more rapidly from the date of PD diagnosis, more specifically how early PD patients developed MCI and dementia after PD diagnosis. Therefore, survival models were constructed with the longitudinal cognition data available, where plasma protein levels could be used as predictors. MoCA < 26 and MoCA < 21 were used as thresholds for PDMCI and PDD, respectively. Not all PDD patients had received a formal PDD diagnosis, but the literature supports these cut-offs as an estimation of cognitive levels at PDD and PDMCI [175], especially in the absence of other factors that could explain the cognitive score.

According to the patient history available, there was no data on the patients having severe cognitive impairment at the time of inclusion. PD patients' MoCA scores at the time of diagnosis were not

available for all patients and it was assumed that they all had similar cognitive function to the general population at baseline. Hence, using a formula for estimating MoCA based on age and sex in a healthy Swedish population [176], a baseline population estimated MoCA score was set at time of diagnosis. A study by Borland et al. suggests that for an estimation of cognitive function, one could use the formula MoCA = 31.104 + 0.565S − 0.090A + 0.713E, where S = sex(0=men,1=women), A = age (years), E = level of education (1 = primary, 2 = secondary, 3 = higher). All time points with available MoCA scores were used to fit a linear regression model for each patient. MoCA score was a function of disease duration and the slope was used as the rate of decline. The formula $MoCA(years) = rate * years + MoCA_{Baseline}$ was used to calculate when the patients would reach MoCA score <26 and <21 respectively. The model was not extrapolated beyond the last available MoCA score, as it was not known how the patients were progressing after that, and the patients were considered lost to follow-up beyond that point. This data was used to fit a cox proportional hazards model for each quantified protein against the PD population. This linear model fit is an estimation of PDMCI and PDD conversion, and real-life cognitive decline can fluctuate as seen in Figure 4.

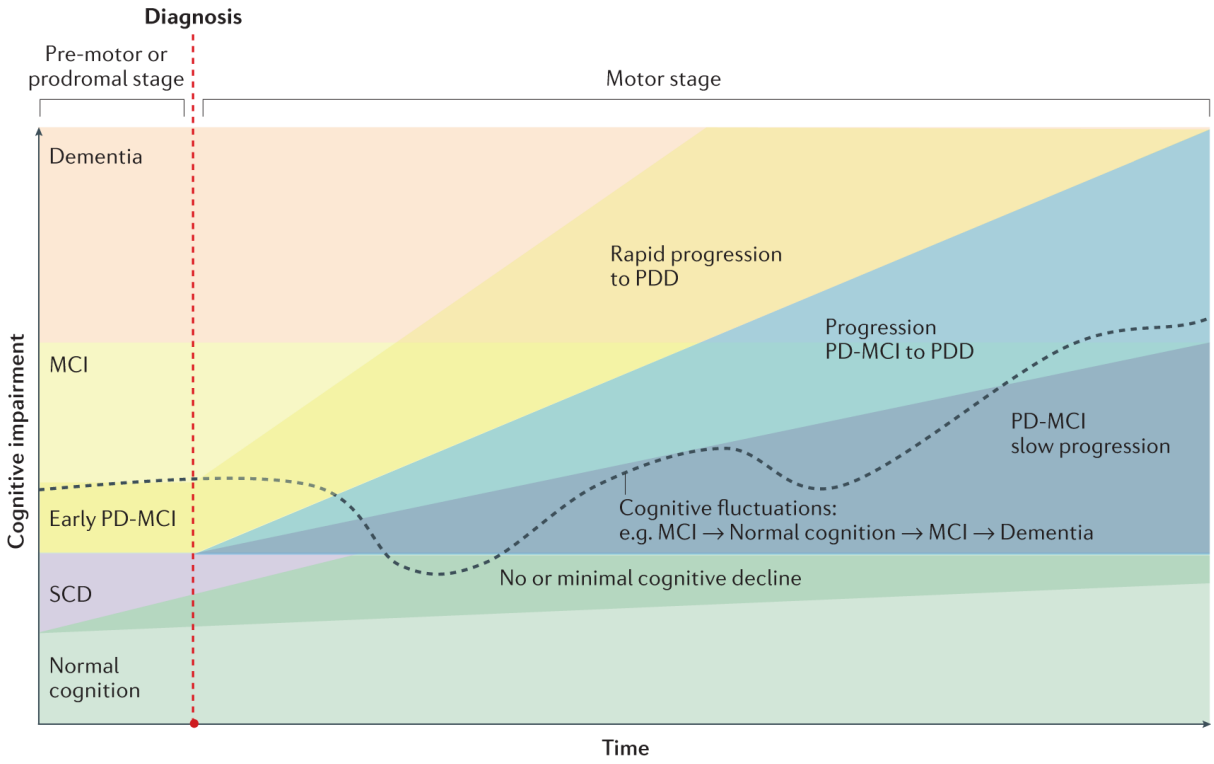

FIGURE 4 – TYPICAL PROGRESSION OF COGNITIVE DECLINE IN PARKINSON'S DISEASE (FIGURE ADAPTED FROM *PARKINSON DISEASE-ASSOCIATED COGNITIVE IMPAIRMENT*, AARSLAND, 2021 [64])

2.1.4 Cohort for complement study

Samples for the complement study were selected from the Biopark cohort. An effect size was estimated off the literature as a rough guide to the sample size. From the difference in mean in complement levels in plasma between controls and neurological disease and their SD (Cohen's d of 0.68) [177, 178], two groups of 40 individuals in each gave a p-value = 0.03 for power = 0.8. As the aim was also to explore clinical PD phenotype with levels of complement factors, the selected PD group was larger than the comparison groups, hence used 81 PD patients and 48 age and sex matched controls (chi squared $p>0.05$ for sex, t-test $p>0.05$ for age). A further 23 individuals were included in the study, clinically diagnosed with 4R- Tauopathies, either Progressive Supranuclear Palsy (PSP) or Corticobasal Syndrome (CBS). This APD group was used to assess the specificity of any findings in PD, but owing to sample availability, this group was smaller and not age and sex matched. PD patients were included if they had a clinical diagnosis of PD set by a neurologist [2]. PSP and CBS diagnoses were set clinically by at least two neurologists, and patients were included if PSP or CBS was the most probable diagnosis[179, 180]. As differential diagnosis between PSP and CBS was sometimes difficult, and due to their pathological proximity and to increase power, they were combined into one APD group (4R-Tauopathies). Subjects from all groups were not selected for the study cohort if they had autoimmune disorders, inflammatory disorders, infections, were taking antibiotics or any form or regular immune-modulating medication.

A further 78 participants (58 PD and 20 HC) where serum was collected were further recruited from the Aetionomy cohort [125] for the CH50 complement activity assay. The inclusion and exclusion criteria were the same as for the Biopark samples.

2.2 Material - kits and chemicals

CH50 test kit (K002, HaemoScan)

Cytiva Immobiline™ DryStrip Gels (IPG strips), 24cm (10697465, Fisher Scientific Ltd)

Dithiothreitol (DTT) (D5545, Merck Sigma-Aldrich)

Glycerol solution (49781, Sigma-Aldrich)

High Select™ Depletion Spin Columns (A36369, Thermo Scientific™)

Human C9 (Complement component C9) ELISA Kit (EH0673, FineTest)

Iodoacetamide (IAA) (I1149, Merck, Sigma-Aldrich)

MILLIPLEX Human Complement Panel 1 (HCMP1MAG-19K, Merck Millipore)

MILLIPLEX Human Complement Panel 2 (HCMP2MAG-19K, Merck Millipore)

OFFGEL IPG-Buffer, pH 3-10 (GE17-6000-87, Merck)

Pierce™ Acetonitrile (ACN), LC-MS Grade (51101, Thermo Scientific™)

Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (84868, Thermo Scientific™)

Pierce™ Peptide Desalting Spin Columns (89852, Thermo Scientific™)

Pierce™ Water, LC-MS Grade (85189, Thermo Scientific™)

ProteoPrep® Immunoaffinity Albumin and IgG Depletion Kit (PROTIA, Sigma-Aldrich)

Sequencing Grade Modified Trypsin (V5117, Promega)

Simoa® NF-light™ Advantage Kit (103186, Quanterix)

Sodium dodecyl sulfate (SDS) 10% (40121008, Bio-world)

SOLA HRP SPE Cartridges (60109-002, Thermo Scientific™)

Thiourea (Part Number: 5188-6436, Agilent)

TMT10plex™ Isobaric Label Reagent Set (90406, Thermo Scientific™)

triethylammonium bicarbonate (TEAB) (T7408, Merck Sigma-Aldrich)

Trifluoroacetic acid (TFA), hypergrade for LC-MS (1.08262.0100, Merk Millipore)

tris(2-carboxyethyl) phosphine (TCEP) (C4706, Merck Sigma-Aldrich)

Urea (U5378, Sigma-Aldrich)

2.3 Methods for mass spectrometry – Discovery

2.3.1 Overview

The plasma samples needed to be optimally processed before analysing on the mass spectrometer to maximise sample resolution and accuracy. The high abundant proteins were depleted to facilitate detection of lower abundant proteins. The samples were then digested to break down proteins to peptides. Samples were labelled with isobaric mass tags and pooled to minimise inter run variation. The samples were finally fractionated to maximise resolution on the mass spectrometer. Finally, the study samples were run on the instrument. A schematic protocol overview is shown in Figure 5.

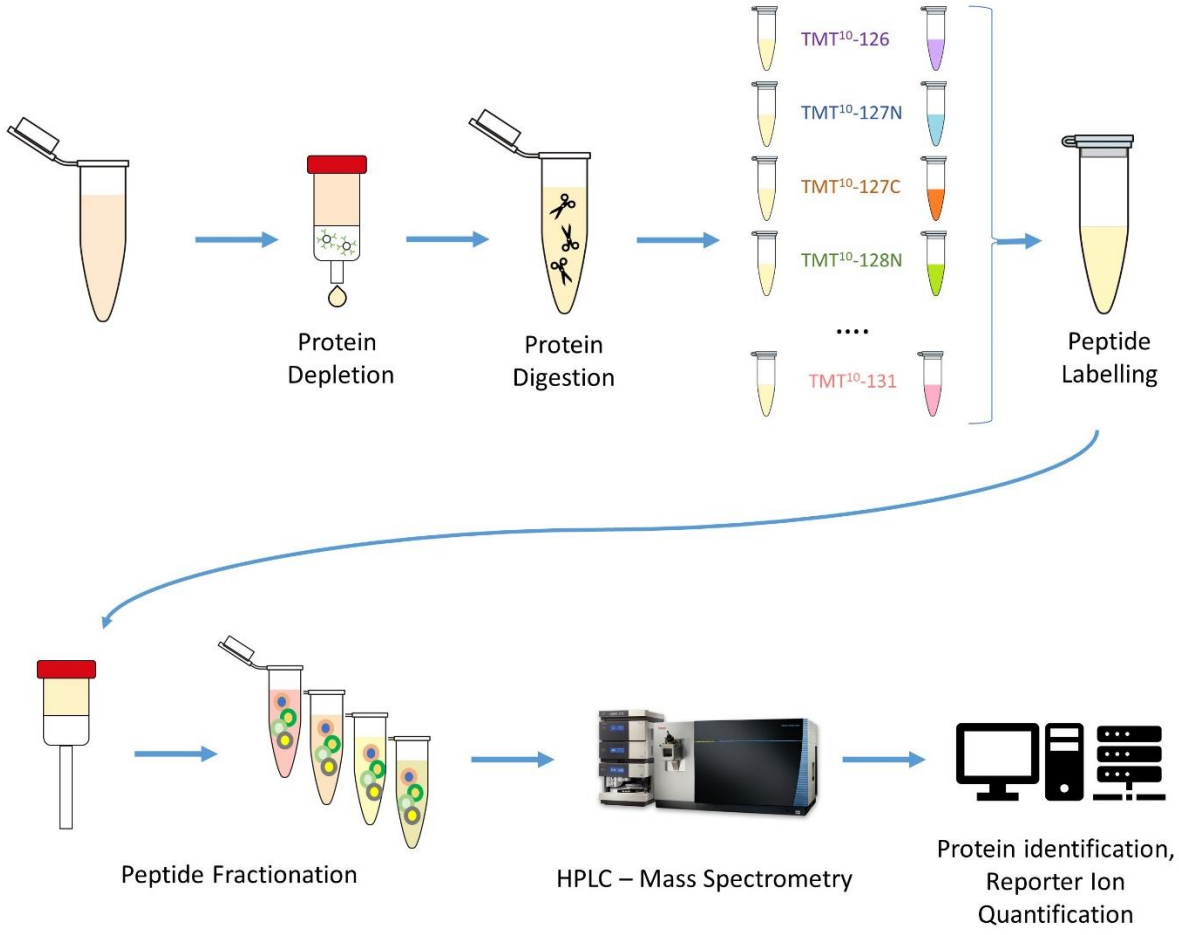

FIGURE 5 - SAMPLE PREPARATION PROTOCOL FOR UNTARGETED MASS SPECTROMETRY

2.3.2 Sample preparation

The protocol had previously been set up and optimised by the group. However, a few additional protein depletion and fractionation methods were tested to further optimise the setup. Two different protein depletion techniques were compared, one which depleted the top 2 most abundant proteins, and one which depleted the top 14 most abundant proteins. The top 2 depletion (ProteoPrep) had

previously been used, but we wanted to compare its depletion efficacy with available top 14 depletion columns. Similarly, two fractionation methods were compared, one using spin columns separating peptides by hydrophobicity, and one using gel strips separating samples by isoelectric point.

2.3.2.1 Protein depletion – ProteoPrep

The ProteoPrep® Immunoaffinity Albumin and IgG Depletion Kit (PROTIA, Sigma-Aldrich) was used for protein depletion in the discovery study. The ProteoPrep Immunoaffinity medium in the prepacked spin columns is a mixture of two beaded mediums containing recombinantly expressed, small single-chain antibody ligands, resulting in low non-specific binding and high capacity. This kit is targeted toward human albumin and IgG, which comprise about 70% of the total proteins in plasma. The protocol was performed according to the manufacturer's instructions, as summarised below.

30ml of Liquid Chromatography Mass Spectrometry (LCMS) grade water was added to a bottle of ProteoPrep Immunoaffinity Equilibration Buffer.

The bottom tip of the spin columns was broken off, and the cap loosened approximately 1 full turn.

The spin columns were centrifuged in a 2ml collection tube at 5,000 g for 5–10 seconds to remove the storage solution.

400µl of the kit Equilibration Buffer was added to the medium in the spin column and centrifuged at 5,000 g for 5–10 seconds. The Equilibration Buffer was discarded, and the spin column was replaced into the same collection tube.

The above equilibration step was repeated twice.

After the final equilibration, each spin column was centrifuged for 30 seconds, the buffer was discarded, and the spin column was placed into a fresh 2ml collection tube.

Each plasma sample was diluted by adding 35µl of plasma to 70 µl Equilibration Buffer.

The diluted plasma sample was added to the 100µl of the 1:2 diluted plasma to the top of the packed medium bed and incubated at room temperature for 10 minutes.

The spin column was centrifuged at 8,000g for 60 seconds.

The eluate in the collection tube was reapplied to the top of the medium bed and incubated for 10 minutes at room temperature.

The spin column was centrifuged in the same collection tube as before at 8,000g for 60 seconds.

The remaining unbound proteins from the spin column were washed by adding 125µl of the Equilibration Buffer to the top of the medium bed and centrifuged at 8,000g for 60 seconds, and the eluate was collected in the same tube.

The protein concentration of each sample was measured on the Thermo Scientific™ NanoDrop™ One spectrophotometer (Protein A280), with equilibration buffer used as blank.

40µg of each sample was transferred to an Eppendorf tube.

10µg of each sample to a Quality Control Mix tube Eppendorf tube.

The samples were then frozen and lyophilised.

2.3.2.2 Protein depletion – Top 14

The High Select Top 14 Abundant Protein Depletion Resin (A36369, Thermo Scientific™) was used as a comparative depletion method to the ProteoPrep in the optimisation experiment. The resin uses highly specific immobilised antibodies for protein removal, providing minimal nonspecific interactions with other proteins. The columns are intended to remove the top 14 most abundant proteins from plasma, which comprise about 95% of the total amount of proteins in plasma. The 14 proteins listed are Albumin, IgG, IgD, IgE, IgG, IgG (light chains), IgM, Alpha-1-acid glycoprotein, Alpha-1-antitrypsin, Alpha-2-macroglobulin, Apolipoprotein A1, Fibrinogen, Haptoglobin, and Transferrin.

The protocol was performed according to the manufacturer's instructions, as summarised below.

Up to 600µg protein was added to each column resin slurry.

The column was capped and inverted several times until homogeneous.

The column was incubated at room temperature with gentle end-over-end mixing for 10 minutes.

The bottom closure was snapped off, and the cap loosened.

The column was placed in a 2ml tube and centrifuged at 1000 g for 2 minutes.

2.3.2.3 Sample digestion

The plasma samples were reduced, alkylated, and digested to obtain peptides.

To each 40µg lyophilised sample 40µL of 100mM triethylammonium bicarbonate (TEAB) with 0.1% Sodium dodecyl sulphate (SDS) was added.

Reduction

2µL of 200mM tris(2-carboxyethyl)phosphine (TCEP) was added to the sample and incubated at 55°C, shaking at 400rpm for 1 hour for reduction.

Alkylation

Immediately before use, iodoacetamide (IAA) was prepared with 100mM TEAB to make 375mM IAA solution. Solution was protected from light.

2µL of the 375mM iodoacetamide to the sample and incubate for 30 minutes protected from light at room temperature to alkylate the samples.

Samples were then immediately frozen and lyophilised.

Digestion

40µg of protein pellets were resuspended with 40µL of 100mM TEAB 0.1% SDS.

1 µg /µl trypsin solution was prepared by adding 100µl of trypsin solution was added to 100µg lyophilised trypsin.

1.6µl (1.6ug) trypsin was added per 40µg of protein and digested overnight at 37°C 400rpm.

2.3.2.4 TMT-labelling

In order to minimise inter run technical variation, the peptide samples were labelled with tandem mass tags (TMT10plex, cat#90110, Thermo Scientific™). TMTs are isobaric mass tags, which are molecules with the same mass tags but result in different reporter ions post fragmentation. Their structural formulas are shown in Figure 6. This allowed for pooling of samples in batches of 10, which reduced overall technical variation, reduced total run time, and allowed for intra run controls. For each 10-plex, the same proportion of PD to HC samples were run (normally 6:3), and the TMT131 was reserved to the quality control (QC) mix that was used as an intra run control. The QC was composed of equal amounts of each of the 198 study samples.

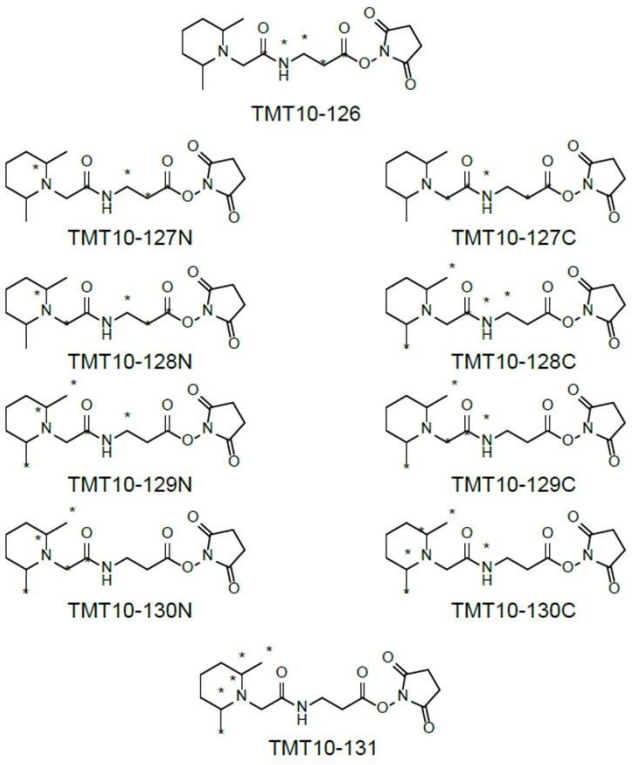

FIGURE 6 – STRUCTURAL FORMULAS OF THE 10-PLEX TANDEM MASS TAGS USED IN THE DISCOVERY STUDY

Immediately before use, TMT Label Reagents were equilibrated to room temperature. For the 5mg vials, 256µl of anhydrous acetonitrile (ACN) was added and let dissolve for 5 minutes.

Carefully 20.5µL of the TMT Label Reagent was added to each 40µL sample (40µg protein digest).

The TMT10-131 was added to the Quality Control Mix.

Mass tags and samples were incubated together for 1 hour at room temperature.

4µL of 5% hydroxylamine was added to the sample and incubated for 15 minutes to quench the reaction.

Equal amounts (approximately 60µl) of each sample were combined in a new microcentrifuge tube (2ml), frozen and lyophilize to dry labelled peptide sample (400µg).

Samples were randomised, and it was ensured a similar proportion of Healthy Controls (HC) and Parkinson's Disease (PD) samples were in each TMT-10-plex, and the TMT-131 was used for the master mix. The sample layout is shown in Figure 7.

TMT	126	127N	127C	128N	128C	129N	129C	130N	130C	131
10 plex 1	HC	PD	PD	PD	PD	PD	PD	HC	HC	QC
10 plex 2	PD	PD	HC	PD	HC	PD	HC	21	PD	QC
10 plex 3	HC	HC	PD	PD	78	PD	PD	PD	HC	QC
10 plex 4	PD	PD	HC	HC	PD	HC	PD	PD	PD	QC
10 plex 5	PD	PD	PD	HC	HC	PD	HC	PD	PD	QC
10 plex 6	HC	PD	HC	HC	PD	PD	PD	PD	PD	QC
10 plex 7	PD	PD	PD	HC	HC	PD	HC	PD	PD	QC
10 plex 8	HC	PD	PD	PD	PD	PD	PD	HC	HC	QC
10 plex 9	HC	PD	PD	HC	PD	PD	HC	PD	PD	QC
10 plex 10	PD	PD	PD	PD	PD	HC	HC	PD	HC	QC
10 plex 11	PD	PD	PD	PD	PD	PD	HC	HC	HC	QC
10 plex 12	PD	PD	PD	HC	HC	PD	PD	HC	PD	QC
10 plex 13	PD	PD	PD	HC	HC	PD	HC	PD	PD	QC
10 plex 14	HC	PD	PD	HC	HC	PD	PD	PD	PD	QC
10 plex 15	PD	PD	HC	PD	HC	PD	HC	PD	PD	QC
10 plex 16	HC	PD	PD	PD	PD	HC	PD	HC	PD	QC
10 plex 17	PD	PD	HC	PD	HC	PD	HC	PD	PD	QC
10 plex 18	PD	PD	HC	HC	PD	PD	PD	PD	HC	QC
10 plex 19	PD	HC	PD	PD	PD	HC	PD	HC	PD	QC
10 plex 20	PD	PD	HC	HC	PD	PD	PD	HC	PD	QC
10 plex 21	HC	PD	PD	HC	HC	PD	237	HC	PD	QC
10 plex 22	HC	HC	PD	PD	PD	PD	23	HC	HC	QC

FIGURE 7 - LAYOUT OF STUDY SAMPLES

Layout of the 198 study samples including Parkinson's Disease (PD), and Healthy Control (HC). The Quality control sample was always labelled with TMT131. This resulted in 22 10-plex sample sets.

2.3.2.5 Peptide fractionation – High pH Spin Columns

Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (cat 84868, Thermo Scientific™) increases protein identification from LC/MS analysis through orthogonal peptide fractionation and was used for the discovery proteomic study.

An equilibration solution was prepared by adding 0.1% Trifluoroacetic acid (TFA) to LCMS grade water.

400µg labelled peptide sample was dissolved in 1.5ml 0.1% TFA solution.

The following fractions of Acetonitrile (ACN) to 0.1% Triethylamine (TEA) solutions were made up for the 10 fractions.

Fraction	% Acetonitrile	% Triethylamine (0.1%)
1	5	100
2	10	90
3	12.5	87.5
4	15	85
5	17.5	82.5
6	20	80
7	22.5	77.5
8	25	75
9	50	50
10	100	0

The protective tip from the bottom of the spin column was discarded and placed into a 2.0ml tube.

The column was centrifuged at 5000 × g for 2 minutes to remove the solution and pack the resin material, and the liquid was discarded.

300µL of ACN was loaded into the column, placed back into the 2.0ml sample tube and centrifuged at 5000 × g for 2 minutes. The ACN was discarded, and the wash step repeated.

The spin column was washed twice with 300µl 0.1% TFA solution at 5000 x g for 2 minutes.

The spin column was placed into a new 2.0ml sample tube and loaded with 300µl of the sample solution onto each column and centrifuged at 3000 × g for 2 minutes.

The column was placed into a new 2.0ml sample tube, loaded with 300µL of water onto the column and centrifuged at 3000 × g for 2 minutes to collect the wash fraction.

The column was placed into a new 2.0ml sample tube. 300µl of the appropriate elution solution (e.g., 5% ACN, 0.1% TEA) and centrifuge at 3000 × g for 2 minutes to collect the fraction. It was repeated for the remaining step gradient fractions using the appropriate elution solutions in new 2.0ml sample tubes.

The samples were subsequently frozen and lyophilised.

2.3.2.6 Peptide fractionation – Gel Strips

OFFGEL (Agilent) Isoelectric Focusing (IEF) separates peptides by their isoelectric point (pI) and is used as a fractionation method prior to mass spectrometry.

OFFGEL stock solution (1.25x) was made up by combining 25.2g urea, 9.1g thiourea, 600mg Dithiothreitol (DTT), 6mL Glycerol, and 0.6mL OFFGEL buffer. The solution was mixed well, and the contents brought to a total volume of 50mL with dH2O.

The fully lyophilised peptide sample was dissolved in 3.6mL of 1x OFFGEL stock solution (0.72mL dH2O + 2.88mL 1.25x OFFGEL stock solution).

The immobilized pH gradient (IPG) strip rehydration solution was prepared. 2.5mL were made up with 0.5mL dH2O + 2.0ml OFFGEL Stock Solution.

The tray was placed in the Agilent 3100 OFFGEL Fractionator with the fixed electrode side to on the left and handle on the right.

The protective backing from the 24cm IPG strip gel, and the strip was placed in the tray with gel side up (side with protective strip that is convex) and the anode/low pH side (+) to the left.

The 24 well frame was attached and secured on top of the gel.

40µL of IPG strip rehydration solution was pipetted into each of the wells.

4 electrode pads were removed for each strip. Using tweezers, one pad was wet in the IPG rehydration solution and placed on the protruding end of the IPG strip, with a second wetted pad placed on top of the first one. The same was repeated on the other end of the strip.

The strip was left to swell for 15 minutes.

150µL of sample was pipetted into each well, and the cover seal was placed on the frame.

10µL of H2O was reapplied onto each electrode pad.

The tray was placed in the instrument platform.

200µL of cover fluid (mineral oil) was applied to the anode end and 700µL to the cathode end. After 1 min additional 200µL was applied to each end of the strip. After 3 min another 100µL was applied to the anode end. The cover fluid should not extend further than ½ the height of the tray grooves.

The fixed anode was placed into the slots on the left side of the tray on top of the electrode pads. The tray was slid into the anode connector.

The movable cathode was inserted on the electrode pads on the right side.

The lid was closed to start the fractionation. The fractionation was done after approximately 48h. After 24h the upper electrode pads were replaced.

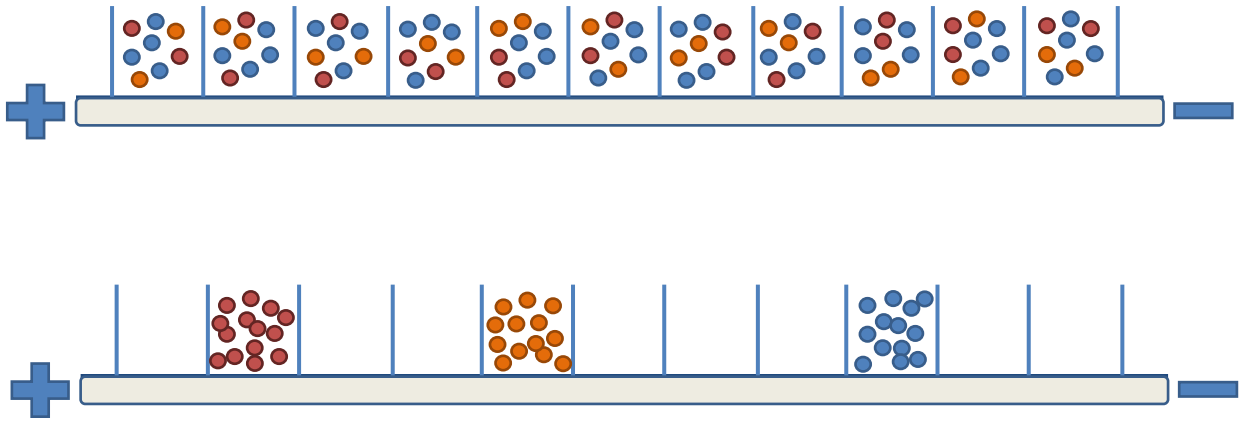

FIGURE 8 – OFFGEL PEPTIDE SEPARATION PRINCIPLE

Overview of OFFGEL isoelectric focusing peptide fractionation. The peptide sample is added in all wells and a voltage is applied across the gel. Peptides move over time to their isoelectric point on the gel.

Once the fractionation was completed, the sample was removed from each well and desalted with solid phase extraction (SPE) SOLA cartridges.

The columns were conditioned with 500µl methanol and equilibrated with 500µl dH2O. the fractionated peptide sample was added to the column, washed with 500µl 5% methanol in dH2O, and eluted with 500µl methanol.

Samples were lyophilised to completion.

2.3.3 Mass spectrometric run

The lyophilised peptide samples were run on the mass spectrometer according to the following protocol by the Centre of Excellence for Mass Spectrometry (CEMS)-Denmark Hill Proteomics Facility.

2.3.3.1 Liquid Chromatography Tandem Mass Spectrometry acquisition

Prior to LC-MS/MS analysis, peptide fractions were reconstituted in 0.05% trifluoroacetic acid (TFA), 2% ACN and de-ionised water then shaken at 37°C, vortexed and centrifuged thoroughly.

Each peptide extract was reconstituted to a calculated concentration of 0.5µg/µL with retrospective protein concentration of peptide extracts estimated by spectrophotometry (NanoDrop™ One, Thermo Scientific™) of the plasma samples.

No	Time	Flow	%B	Curve
1	0.000	Run		
2	0.000	0.250	5.0	5
3	5.000	0.250	10.0	5
4	80.000	0.250	35.0	5
5	95.000	0.250	90.0	5
6	95.100	0.250	90.0	5
7	100.000	0.250	90.0	5
8	100.100	0.250	5.0	5
9	120.000	0.250	5.0	5
10	*New Row*			
11	70.000	Stop Run		

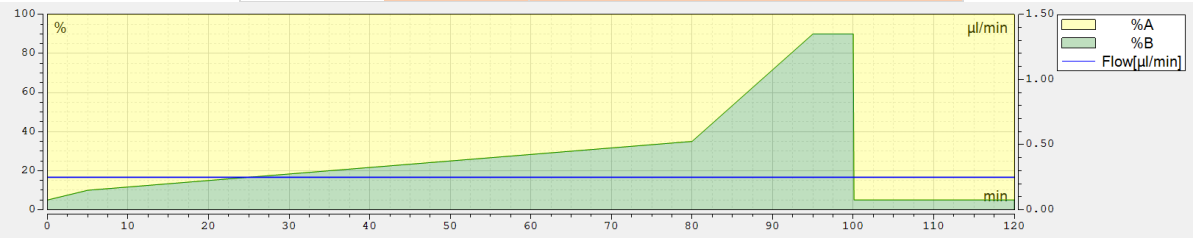

FIGURE 9 – METHOD OF CHROMATOGRAPHIC SEPARATION FOR TMT10PLEX

A linear increase of 5% solvent B (0.1% formic acid in 80% acetonitrile) was run for 120 minutes before a 5-minute wash at 95% solvent B.

Chromatographic separation was achieved by a two-column configuration; 4µL of sample (~2µg) was injected first onto a nano-trap column (Acclaim PepMap100 C18 Trap, 5 mm x 300 µm, Thermo Scientific) packed with octadecyl carbon chain C18-bonded silica (C18) using the Thermo Scientific nanoflow LC system UltiMate 3000 RSLC nanosystem. Peptides were then resolved using a linear gradient of 0.1% FA in 80% ACN (10% to 65% over 120 minutes) through a nanocolumn (EASY-Spray PepMap® RSLC C18, 2µm 100 Å, 75µm x 50cm), set at 40°C and connected to an EASY-Spray ion source (Thermo Scientific) at a flow rate of 250nL/min (Figure 9). Mass spectra were acquired on a Thermo Scientific Orbitrap Fusion Lumos instrument throughout the chromatographic run which operated in data-dependent mode to automatically switch between full scan MS and MS/MS acquisition. Instrument control was through Tune 3.4.0 and Xcalibur 4.2 (Thermo Scientific).

The instrument was programmed to using a Synchronous Precursor Selection with Multinotch MS^3 method (SPS). Synchronous Precursor Selection is a process of selecting multiple MS2 precursors using a single fill and single waveform in a collision-induced dissociation (CID) or Higher collision-induced dissociation (HCD) cell, while Multinotch MS3 is to reduce co-isolated interference from MS2 in an

ion-trap cell. This method allows for accurate and sensitive quantitation based on isobaric TMT tags (Figure 10).

The method was set to a 3 second cycle time between a full Orbitrap MS scan, Ion Trap MS/MS fragmentation and finally HCD Orbitrap scan for MS3 fragments. Orbitrap spectra (FTMS1) were collected at a resolution of 120,000 over a scan range of m/z 375-1500 with an automatic gain control (AGC) setting of 4.0e5 and maximum injection time of 35 ms. Monoisotopic precursor ions were filtered using charge state (+2 to +7) with an intensity threshold set between 5.0e3 to 1.0e20 and a dynamic exclusion window of 35 secs ± 10 ppm. MS2 precursor ions were isolated in the quadrupole set to a mass width filter of 0.7 m/z. Ion trap fragmentation spectra (ITMS2) were collected with an AGC target setting of 1.0e4 with a maximum injection time of 35 ms with CID collision energy set at 35%. Ions detected with isobaric tags are further fragmented under HCD at 65% collision energy at a resolution of 60,000 over a scan range m/z 100-500. The AGC target was set at 5.0e4. The Top 5 SPS precursors were selected for fragmentation for each detected ion.

FIGURE 10 – SYNCHRONOUS PRECURSOR SELECTION MS3 METHOD

2.3.3.2 Pre-processing of raw LC-MS/MS data

Raw data files produced in Xcalibur software 4.2 were processed using Proteome Discoverer, (ver. 2.3; Thermo Scientific) to determine peptide identification (Figure 11); the subsequent Mascot (ver. 2.6; available at: http://www.matrixscience.com) and Sequest (Eng *et al.* [181]) output file was used for additional pre-processing and analysis.

Prior to database searching a spectrum selector threshold was applied; minimum 700 Da and maximum 10000 Da. Within Mascot and Sequest, mass spectra were searched against the

uniprot/Swiss-prot database (ver. 2021, 50,830 entries), taxonomy was set to human with precursor and fragment mass tolerances were set to ±10ppm and 0.6 Da, respectively. Dynamic modifications included TMT6plex modification of the lysine residue (TMT (K)), TMT6plex modification of the peptide N-Terminus (TMT (N-Term)), and oxidation of the methionines. Additionally, carbamidomethylation of cysteine residues was set as a fixed modification and two miss cleavages were allowed. Validation of merged Mascot and Sequest database results were conducted by Percolator at FDR of 0.05.

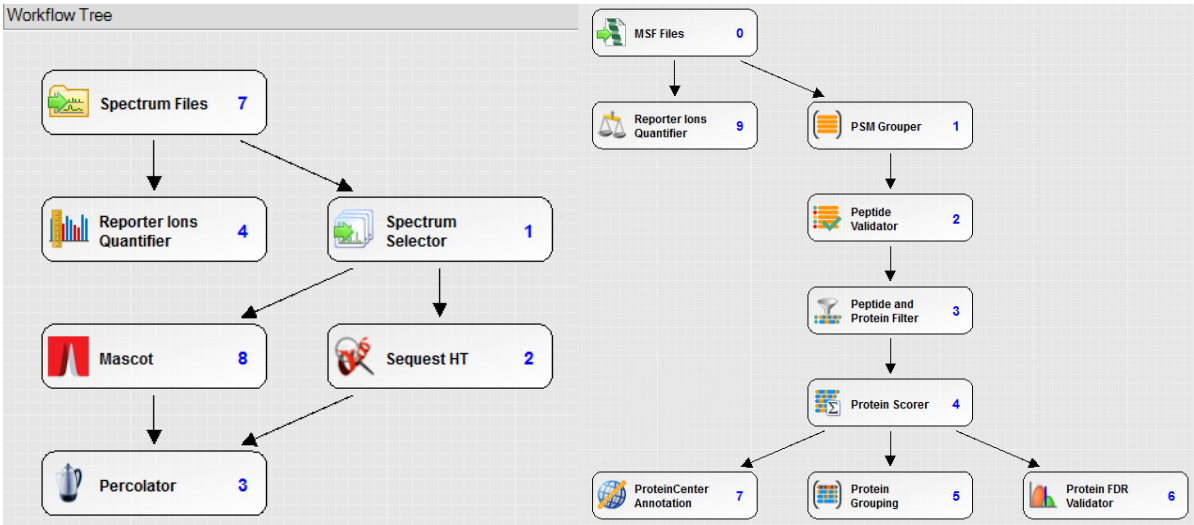

FIGURE 11 – PROTEOME DISCOVERER NODAL WORKFLOW FOR RAW DATA PROCESSING FOR DATABASE SEARCHING AND CONSENSUS QUANTIFIER ANNOTATION METHOD.

Proteome Discoverer workflow for database searching of MS/MS spectra in Chapter 3. Spectrum selector thresholds were set as 700Da-10000 Da. In Mascot and Sequest, precursor and fragment tolerances were set to ±10ppm and 0.6 Da. Modifications included TMT (K), TMT (N-Term), oxidation (M) and carbamidomethylation (C) of cysteine residues. Two miss cleavages were allowed.

2.3.4 Data processing

2.3.4.1 Preprocessing

Total protein distributions per sample were visualised in Proteome Discoverer, (ver. 2.3; Thermo Scientific). Normalisation to total protein abundance was also performed in Proteome Discoverer, and batch effects due to group, sample, and mass tag were inspected. The normalised data was exported and analysed in RStudio (R version 4.2.1).

An expression set was constructed in R using the limma package with protein features as feature data, clinical characteristics as phenotype data, and protein abundances as matrix data. Plots were made

using the ggplot2 package. Density plots for arithmetic and log10 transformed data was visualised for each sample and each protein. Shapiro-wilk p-value was calculated for each protein distribution to assess if most proteins are normally distributed. Coefficients of Variation for proteins across all samples and across QCs were calculated, and distributions visualised. QCs were corrected for by subtracting the QC quantity from the study sample quantity for each protein in each 10-plex and added to the average protein abundance for each protein. The limma modified PCA plot was used to find technical batch effects and corrected for using the "removeBatchEffect" function.

Proteins that were quantified in less than half the subjects were removed. Correlation scatterplots and boxplots were visualised in ggplot2 with parametric statistics for proteins known to vary with demographic parameters.

2.3.4.2 Statistical analysis

For dimensionality reduced analysis (Principal component analysis (PCA) and t-distributed stochastic neighbour embedding (t-SNE)), the residual matrix was used after correcting for age and sex with a linear model. Missing values were imputed with the missMDA package. The factoextra package with the fviz_pca function was used to create PCA plots, and the data was centred and scaled. T-SNE plots were constructed with the Rtsne package with the data reduced to two dimensions.

Linear models for the proteomic data were analysed using the limma package which was originally designed for microarray data. An interception-free model matrix with variable of interest (i.e. disease group), and covariates (i.e. age and sex) were fit in a model matrix against the proteome. The variable contrast of interest (i.e. PD versus HC) was defines, and a robust linear model was fit to the data. The fitted model object was processed by a contrast matrix before being passed to eBayes to obtain the statistical output. P-values were adjusted by the Benjamini-Hochberg method. The EnhancedVolcano package was used to generate volcano plots.

The machine learning models for diagnostic accuracy were constructed for the most significant proteins using the caret package. Data was partitioned 5 times into train and test sets using a 90/10 split. The training data was used to train an effective model using leave one out cross validation (LOOCV), both with and without covariates as "train(Group~., data=Training_data, method="glm", trControl=trainControl(method = "LOOCV"))". Importance of variables in the model were calculated. The model was used to predict the test data. Both accuracy and area under curve (AUC) were used as outcome measures.

Support vector machines (SVM) were used to build biomarker panels. The e1071 was used for SVMs, and custom codes for multiple recursive feature elimination SVM were implemented("mSVM-RFE" by

johncolby GitHub, based on algorithm by Duan et al. [182]). For each loop, 4 iterations of a 5-fold cross validation of the SVM-RFE algorithm were run, and the outcome model was used to predict diagnosis accuracy, and area under curve for a receiver operating characteristic (ROC) curve was calculated. For the next loop, the lowest ranked protein was eliminated, and the SVM reconstructed. This loop was reiterated until one protein was left.

For pathway analysis, the enrichR package was used, and the top protein hits were searched in the KEGG 2019 Human database [183], and p-values for probability of pathway involvement were calculated.

STRING functional protein-protein interactions were investigated using the STRINGdb package. STRING v.11.5 [184] was used, with species 9606 (Homo Sapiens) selected and score threshold of 200.

Correlations with MoCA score were calculated with robust linear models with the limma package similar to the group comparisons. Age, sex and disease duration were used as covariates.

Survival analysis was performed with Cox proportional hazard regressions. Statistics were calculated with the survival package and plotted with the survminer package. Survival probability and hazard ratio for conversion to mild cognitive impairment and dementia were calculated.

For everything else, parametric tests (t-test and ANOVA) used for group comparisons, and Pearson r for correlations. For all linear models, covariates including age, sex, and disease duration were used where appropriate and without accounting for interactions. For plotted age and sex adjusted data, residuals from a linear model were used.

2.4 Methods for mass spectrometry – Verification

2.4.1 Overview

Plasma samples were processed before analysing on the mass spectrometer. Neat plasma from the study participants was reduced, alkylated, trypsinated, and desalted. A parallel reaction monitoring (PRM) method was optimised using a sample mix, and transitions for quality peptides were preselected. Finally, the study samples were run on the instrument. The methods summary is shown in Figure 12.

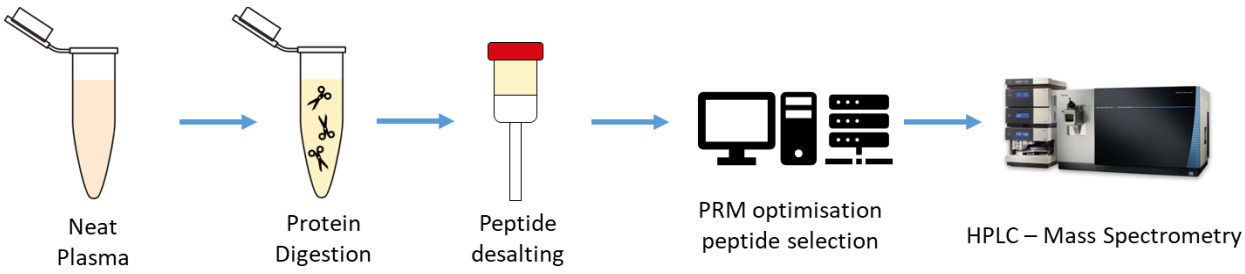

FIGURE 12 - SUMMARY OF METHOD FOR VERIFICATION STUDY

2.4.2 Sample preparation

Plasma concentration was determined on the NanoDrop. 800µg of plasma was aliquoted per sample.

Urea was added to the samples to improve solubility and denaturation of proteins. 8M urea solution was prepared by adding 4.8g of urea powder (MW 60.06) to 7mL of water, dissolve and adjust volume to 10mL, to make 8M urea. 80µl of 8M urea was added to the plasma (which was on average 20µl).

Reduction

100mM DTT (MW 154.25) was used for reduction. 5µl DTT was added to the plasma and incubated at 56C for 1h.

Alkylation

Immediately before use, iodoacetamide (MW 184.96) was dissolved in water to make 300mM iodoacetamide. 5µl IAA was added to the plasma, and incubated in the dark, at room temperature, for 1 hour. The reaction was quenched by adding 10µl 100mM DTT solution.

Digestion

100mM 0.1% SDS TEAB was made up by mixing 1M TEAB with water 10% SDS solution.

720µl 100mM TEAB was added to plasma samples to dilute samples to around 1µg/µl, and to dilute the urea to about 1M.

1µg/µl trypsin solution was prepared.

20µl (20ug) of trypsin was added to the sample, and let digest overnight at 37°C 400rpm.

Samples were then lyophilised. Freeze and lyophilise

Desalting

0.1% TFA solution in water was made up as the wash solution. 0.1% TFA in 50% water and 50% ACN was made up as the elution solution

Peptides samples were dissolved in 300µl 0.1%TFA in water.

The tip was removed from the Pierce™ Peptide Desalting Spin Columns, placed in 2ml tube, centrifuged at 5000 g for 1 min.

300µl ACN was loaded in the column and centrifuged at 3000 g for 1 min. This conditioning step was repeated once.

The column was washed by adding 300µl 0.1% TFA in water and centrifuged at 3000g for 1min. This washing step was repeated once.

300µl of sample was added to the column and centrifuged at 3000g for 1min.

The column was washed 0.1%TFA in water by adding 300µl and centrifuged at 3000g for 1min. This step was repeat 2 times.

The samples were eluted with 300µl 0.1% TFA in 50% water 50% ACN. The columns were centrifuged at 3000g for 1min in a new 2ml collection tube.

Additional 300µl 0.1% TFA in 50% water 50% ACN was added to the column and centrifuged at 3000 g for 1 min. The solution was collected int the same tube.

75µl (100ug) of the sample was aliquoted, frozen, and lyophilised.

2.4.3 Mass spectrometric analysis

The mass spectrometric analysis was performed by facility managers at the Centre of Excellence for Mass Spectrometry (CEMS)-Denmark Hill Proteomics Facility.

The top 70 candidate proteins for PD diagnosis and PD cognition were assessed to determine whether they had high quality peptides for the PRM analysis. Peptides were selected if they were unique to the

protein, within the mass range of the instrument, fully tryptic, not susceptible to amino acid modifications, and had doubly or triply charged precursor ions [138]. A plasma mix of 8 PD and 8 HC samples was analysed on the mass spectrometer and the high-quality peptides were analysed. The signal intensity, reproducibility, and chromatographic peak shapes were assessed. In the end 22 peptides from 12 proteins were of high enough quality (Table 2).

TABLE 2 – SHORTLIST OF QUALITY PEPTIDES THAT WERE VERIFIED BY PARALLEL REACTION MONITORING

Peptide	Protein	Peptide Retention Time	Precursor Charge	Precursor Mz
DALSSVQESQVAQQAR	sp|P02656|APOC3_HUMAN	28.08	2	858.929196
TEHYEEQIEAFK	sp|P02748|CO9_HUMAN	34.03	3	508.571939
DVVLTTTFVDDIK	sp|P02748|CO9_HUMAN	32.86	2	733.392864
AIEDYINEFSVR	sp|P02748|CO9_HUMAN	33.22	2	728.359356
EAVEHLQK	sp|P06727|APOA4_HUMAN	32.86	2	477.256174
LEPYADQLR	sp|P06727|APOA4_HUMAN	31.39	2	552.787838
IDQNVEELK	sp|P06727|APOA4_HUMAN	30.33	2	544.285128
IDQTVEELR	sp|P06727|APOA4_HUMAN	30.7	2	551.790578
ISASAEELR	sp|P06727|APOA4_HUMAN	30.7	2	488.258913
SQDILLSVENTVIYR	sp|P07225|PROS_HUMAN	34.99	2	875.472708
NNLELSTPLK	sp|P07225|PROS_HUMAN	31.36	2	564.816595
GTLLALER	sp|P07996|TSP1_HUMAN	32.43	2	436.763634
GPGGVWAAEAISDAR	sp|P0DJI8|SAA1_HUMAN	32.76	2	728.862597
FFGHGAEDSLADQAANEWGR	sp|P0DJI8|SAA1_HUMAN	37.64	3	726.659358
AVYEAVLR	sp|P12955|PEPD_HUMAN	34.12	2	460.763634
VYALPEDLVEVNPK	sp|P13796|PLSL_HUMAN	41.17	3	529.287251
AEIEYLEK	sp|P14151|LYAM1_HUMAN	30.89	2	497.758215
SLEDLQLTHNK	sp|P51884|LUM_HUMAN	34.56	3	433.229736
ILGPLSYSK	sp|P51884|LUM_HUMAN	32.8	2	489.286942
QHGPNVCAVQK	sp|Q15582|BGH3_HUMAN	32.18	3	432.215544
VLTDELK	sp|Q15582|BGH3_HUMAN	26.97	2	409.236918
GAIENLLAK	sp|Q9UPN9|TRI33_HUMAN	30.83	2	464.776742

Two sample injections for all 198 samples were run on the mass spectrometer. Quality controls samples (made up of equal amounts of all 198 samples) and blanks were run after every 10 samples. Desalted peptides were resuspended in 0.1% formic acid and 2% acetonitrile (ACN). All MS analyses were carried out using Orbitrap Lumos mass spectrometer (Thermo Fisher Scientific) coupled with an UltiMate 3000 RSLC nano ultra-high-performance liquid chromatography (HPLC) system (Thermo Fisher Scientific). Briefly, peptides were loaded and separated on an Easy-spray C18 nano column (15 cm length, 75 mm internal diameter; ThermoFisherScientific) within a nanoflow HPLC (RSLC Ultimate 3000) coupled on-line to a nano-electrospray ionization Orbitrap Fusion Lumos mass spectrometer (ThermoFisherScientific). Peptides were eluted with a linear gradient of 4%–60% buffer B (80% ACN, 0.1% formic acid) at a flow rate of 750 nl/min over 70 min at 45°C. HPLC elution gradient shown in Figure 13.

No	Time	Flow	%B	Curve
1	0.000		Run	
2	0.000	0.750	4.0	5
3	5.000	0.750	5.0	5
4	40.000	0.750	40.0	5
5	50.000	0.750	60.0	5
6	50.100	1.000	90.0	5
7	55.000	1.000	90.0	5
8	55.100	1.000	4.0	5
9	70.000	1.000	4.0	5
10	*New Row*			
11	70.000		Stop Run	

FIGURE 13 - HPLC ELUTION GRADIENT FOR USED FOR PARALLEL REACTION MONITORING

A linear increase of 5% solvent B (0.1% formic acid in 80% acetonitrile) was run for 120 minutes before a 5-minute wash at 95% solvent B.

The instrument was programmed within Xcalibur 4.3 to acquire MS data using "Parallel Reaction Monitoring" method (PRM). PRM is a process of selecting predefined MS1 precursors (m/z and rt) using a single fill and single waveform in a HCD cell, followed by fragmenting them into product ions at MS2 level. All setting parameters were listed in Figure 14.

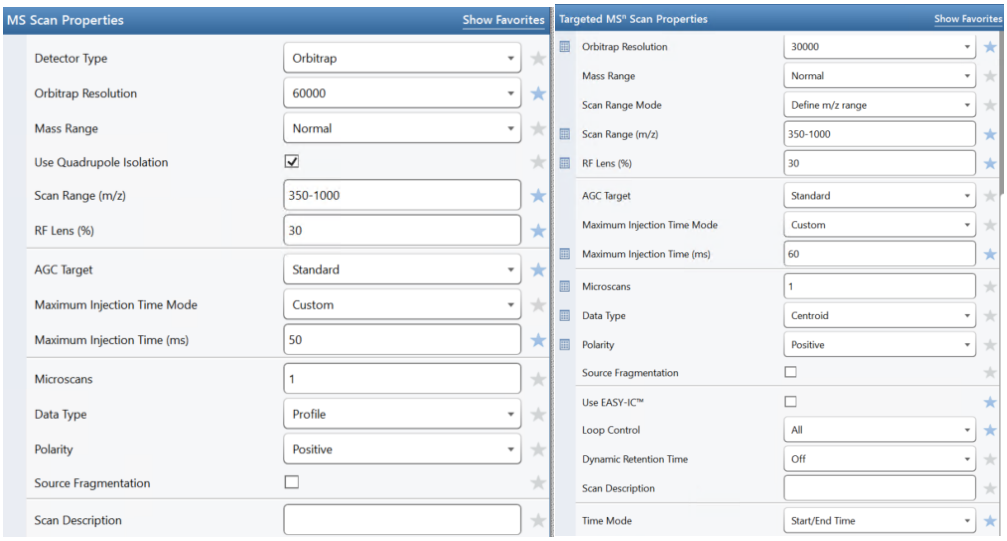

FIGURE 14- SETTINGS USED ON XCALIBUR FOR MASS SPECTROMETRY DATA ACQUISITION

2.4.4 Data processing

2.4.4.1 Data pre-processing

LCMS .raw files processed within Skyline software. After LCMS acquisition, .raw files for 198 samples (double injections each) plus quality controls and blanks are generated and used for the following analysis. All .raw files were loaded into Skyline for data processing, including peak picking and peak integration.

2.4.4.2 Statistical analysis

Peak areas for all peptides for all samples were exported from Skyline and processed in RStudio (R version 4.2.1). Data frames were cleaned up using the tidyr package. The corrplot package was used for correlation plots, ggplot2 was used for all the other plots.

Protein distributions were visually inspected with density plots, and log10 transformed when heavily right skewed. Technical variation and batch effects were visually assessed with scatter plots. Run order, study group, blanks, quality controls (QC), retention times, technical batches, and total ion count were all assessed.

Total ion count was corrected for by dividing the arithmetic protein abundance with the total ion count. QC was adjusted for by fitting a local polynomial regression (loess) function with a span of 0.2 as “stat_smooth(method = "loess", span=0.2)”. A theoretical QC was subsequently predicted for each data point through “predict(loess(y~x, data = data, span = 0.2)”. The predicted QC value was then subtracted from each data point.

Peptide levels were compared with proteins levels from the discovery with correlation plots. Pearson correlation coefficient with significance (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001) were calculated and displayed. Discovery study expression set was read with the limma package.

Average peptide level was defined as the average peptide quantity of replicates 1 and 2. Average protein level was defined as the average of all peptides level averages for each protein.

Parametric tests (t-test and ANOVA) used for group comparisons, and Pearson r for correlations, to make results comparable with the discovery analysis.

Survival analysis was performed with Cox proportional hazard regressions. Statistics were calculated with the survival package and plotted with the survminer package. Survival probability and hazard ratio for conversion to mild cognitive impairment and dementia were calculated.

For all linear models, covariates including age, sex, and disease duration were used where appropriate and without accounting for interactions. For plotted age and sex adjusted data, residuals from a linear model were used.

2.5 Neurofilament light chain assay

2.5.1 Assay protocol

The neurofilament light chain assay (Simoa® NF-light™ Advantage Kit, cat#103186) was run on the Single molecule array (SIMOA). The SIMOA is a bead based ultra-sensitive immunoassay capable of accurately quantifying proteins below femtomolar levels in biofluids [185]. The assay was analysed on the fully automated HD-1 instrument according to manufacturer's instructions. The same set of samples were used as in the mass spectrometry experiments.

Calibrators and samples and reagents were equilibrated to room temperature.

The NF-light assay definition for the HD-1 instrument was downloaded (http://portal.quanterix.com) and imported into the Simoa software.

Eight calibrators ranging from 0-450 pg/ml were provided in the kit. Calibrators were run in duplicates by pipetting 200µl in the first 2 columns of a 96-well SIMOA plate.

Plasma samples were diluted 4x using the Sample Diluent provided. 200µl was pipetted in each well, samples were run in single replicates.

The HD-1 instrument was initialised.

Bead Reagent, Detector Reagent, and SBG (streptavidin-β-galactosidase) Reagent were put into a reagent rack with the barcodes visible and scanned in. The beads were vortexed for 30 seconds immediately prior loading on the instrument.

RGP (Resorufin-D-galactopyranoside) bottle was loaded into the RGP rack and load into the selected lane in the sample bay.

A plate layout was setup on the instrument, calibrators and samples were selected and all run using the "Neat" protocol.

After the run a 4-parameter logistic curve fit, 1/y2 weighted was generated, example of a NfL standard curve is shown in Figure 15.

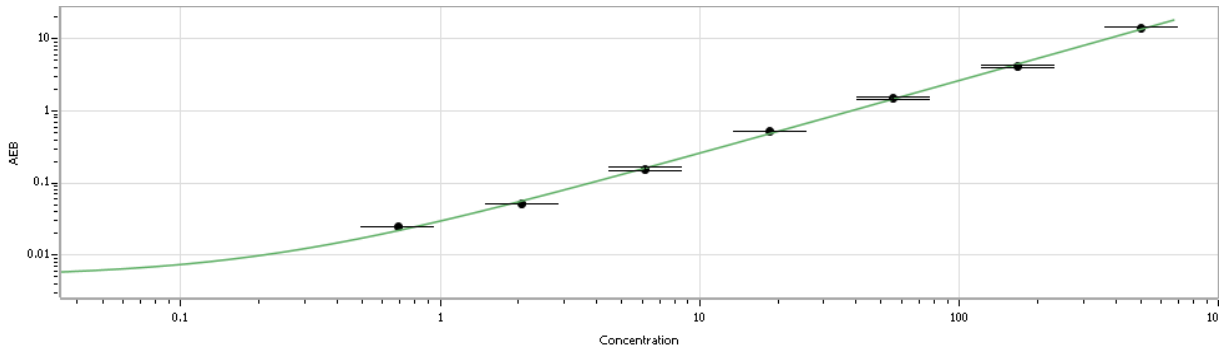

FIGURE 15 – 4-PARAMETER LOGISTIC STANDARD CURVE FIT, EXAMPLE FROM NEUROFILAMENT LIGHT CHAIN ASSAY

2.5.2 Statistical analysis

As samples were diluted offline, the plasma concentrations were calculated by multiplying the results with the dilution factor.

Neurofilament light chain (NfL) concentrations were log10 transformed to achieve normal distribution, which was assessed with density plots and histograms. The data was subsequently corrected for inter-plate variation using a linear model before analysis.

Group level analyses were performed using linear models with Diagnosis as the explanatory variable, both with and without age and sex as additional explanatory variables. The PD diagnosis was further divided into PDND (PD with no dementia) where MoCA scores were ≥ 26, and PDCI (PD with cognitive impairment) where MoCA <26. Spearman's rank was used for correlations. Figures with age and sex adjusted data used residual values from a linear model with age and sex as explanatory variables.

Cox regressions and survival plots were constructed like for the discovery study.

All analysis was performed in RStudio (R version 4.2.1). Plots were made using the ggplot2, ggpubr, and ggbeeswarm packages.

2.6 Complement study

2.6.1 Complement factor quantification

Plasma was chosen for quantification of complement proteins, as the added EDTA prevents further complement activity *in vitro* after blood draw [186]. Complement factors were measured with 3 different assays, two on the Luminex platform, and one sandwich ELISA. Complement factors, products and regulators of the complement system were measured using MILLIPLEX Human Complement Panel 1 (complement C2, C4b, C5, C5a, Factors I and D, and Mannose Binding Lectin; cat# HCMP1MAG-19K), Human Complement Panel 2 (complement C1q, C3, C3b, C4, Factors B and H; cat# HCMP2MAG-19K), and Human C9 (Complement component C9) ELISA Kit, by Fine Test cat# EH0673. Samples and standards were run in duplicates for all assays.

2.6.1.2 Milliplex Complement Panels 1 and 2

PREPARATION OF SAMPLES AND REAGENTS

TABLE 3 – COMPLEMENT PANEL 1 (TOP) AND PANEL 2 (BOTTOM) ANTIBODY-IMMOBILIZED MAGNETIC BEADS

Complement Panel	Bead/Analyte	Luminex® Magnetic Bead Region	Cat. #
1	Anti-Complement C2 Bead	19	HCC2-MAG
1	Anti-Complement C4b Bead	43	HCC4B-MAG
1	Anti-Complement C5 Bead	44	HCC5-MAG
1	Anti-Complement C5a Bead	45	HCC5A-MAG
1	Anti-Complement Factor D (Adipsin) Bead	73	HADPSN-MAG
1	Anti-Mannose-Binding Lectin (MBL) Bead	74	HMLB-MAG
1	Anti-Complement Factor I Bead	76	HCFI-MAG
2	Anti-Complement C1q Bead	14	HCC1Q-MAG
2	Anti-Complement C3 Bead	34	HCC3-MAG
2	Anti-Complement C3b/iC3b Bead	37	HCC3B-MAG
2	Anti-Complement C4 Bead	39	HCMPC4-MAG
2	Anti-Complement Factor B Bead	51	HCFB-MAG
2	Anti-Complement Factor H Bead	75	HCMPCFH-MAG

Samples were thawed and all assay reagents reached room temperature before use.

Samples were diluted to a final plasma dilution of 1:200 for panel 1, and 1:40 000 for panel 2 using the provided Assay Buffer.

Each antibody-bead vial was sonicated for 30 seconds, followed by 1 minute vortex. 150 µL from each antibody-bead vial was added to the Mixing Bottle and brought to a final volume of 3.0ml with provided Bead Diluent.

Quality Control 1 and 2 were reconstituted with 250µl deionized water and vortexed.

Prior to use, the Complement Panel Standard were reconstituted with 250 µL deionized water and vortexed. This was labelled as Standard 7 and used as the highest concentration on the standard curve. Rest of the standards down to standard 1 were serially diluted 1:2 by adding 150µl of a higher standard to 150 µL Assay buffer. Pure Assay Buffer was used as a blank.

ASSAY PROCEDURE

200 µL of Wash Buffer was added into each well of the plate. The plate was sealed and put on a plate shaker for 10 minutes at room temperature (20-25°C), before decanting the wash buffer.

25 µL of each Standard or Control was added into the appropriate wells. Assay Buffer was used for blank (background).

25 µL of Assay Buffer was added to all wells.

25 µL of plasma Sample (1:100 diluted plasma) or quality control sample was added into the appropriate wells.

The mixing bottle containing bead was thoroughly vortexed, and 25µl of the Mixed Beads added to each well. The bead bottle was vortexed intermittently to avoid settling. This step was performed with lights switched off.

The plates were sealed with a plate sealer, wrapped in aluminium foil, and incubated with agitation on a plate shaker (500-800 rpm) overnight at 2-8°C.

Well contents were gently removed, and the plate washed 3 times as follows. The plate was rested on a plate magnet for 60 seconds to allow complete settling of magnetic beads. The well contents were emptied by gently decanting the plate in an appropriate waste receptacle and gently tapping on absorbent pads to remove residual liquid. The plate was washed with 200 µL of Wash Buffer by removing plate from magnet, adding Wash Buffer, shaking for 30 seconds on a plate shaker, reattaching plate to magnet, letting beads settle for 60 seconds and removing well contents as previously described after each wash.

50 µL of Detection Antibodies were added into each well.

The plates were sealed and covered with aluminium foil and incubated with agitation on a plate shaker for 1 hr at room temperature.

50 µL of Streptavidin-Phycoerythrin was added to each well containing the 50µl of Detection Antibodies.

The plates were sealed, covered with aluminium foil, and incubated with agitation on a plate shaker for 30 minutes at room temperature.

Contents of the wells were then gently removed and washed 3 times as previously described.

170 µL of Sheath Fluid was added to all wells. The beads were resuspended on a plate shaker for 5 minutes.

Samples were assayed and quantified on the Bio-Rad Bio-Plex 200 system. Default gate and bead count settings were used for magnetic beads.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Standard 1	Standard 1	Control 1	Control 1	Plasma 7	Plasma 7	Plasma 15	Plasma 15	Plasma 23	Plasma 23	Plasma 31	Plasma 31
B	Standard 2	Standard 2	Control 2	Control 2	Plasma 8	Plasma 8	Plasma 16	Plasma 16	Plasma 24	Plasma 24	Plasma 32	Plasma 32
C	Standard 3	Standard 3	Plasma 1	Plasma 1	Plasma 9	Plasma 9	Plasma 17	Plasma 17	Plasma 25	Plasma 25	Plasma 33	Plasma 33
D	Standard 4	Standard 4	Plasma 2	Plasma 2	Plasma 10	Plasma 10	Plasma 18	Plasma 18	Plasma 26	Plasma 26	Plasma 34	Plasma 34
E	Standard 5	Standard 5	Plasma 3	Plasma 3	Plasma 11	Plasma 11	Plasma 19	Plasma 19	Plasma 27	Plasma 27	Plasma 35	Plasma 35
F	Standard 6	Standard 6	Plasma 4	Plasma 4	Plasma 12	Plasma 12	Plasma 20	Plasma 20	Plasma 28	Plasma 28	Plasma 36	Plasma 36
G	Standard 7	Standard 7	Plasma 5	Plasma 5	Plasma 13	Plasma 13	Plasma 21	Plasma 21	Plasma 29	Plasma 29	Plasma 37	Plasma 37
H	Blank	Blank	Plasma 6	Plasma 6	Plasma 14	Plasma 14	Plasma 22	Plasma 22	Plasma 30	Plasma 30	Plasma 38	Plasma 38

FIGURE 16 – PLATE LAYOUT EXAMPLE ELISA

2.6.1.3 C9 ELISA

C9 was quantified using the Human C9 Elisa Kit (cat #EH0673).

Samples were thawed and all assay reagents reached room temperature before use.

Samples were diluted to a final plasma dilution of 1:2000 using the provided Sample Dilution Buffer.

Prior to use, the Complement Panel Standard were reconstituted with 1000 µL Sample Dilution Buffer and vortexed. This was labelled as Standard 1 and used as the highest concentration on the standard curve. Rest of the standards down to standard 7 were serially diluted 1:2 by adding 300 µL of a higher standard to 300 µL Sample Dilution Buffer. Pure Sample Dilution Buffer was used as a blank.

10ml Biotin-labelled detection antibody per plate was made up, by diluting 100µl antibody stock solution with Antibody Dilution Buffer at 1:100 and mixed.

10ml HRP-Streptavidin Conjugate (SABC) per plate was made up, by diluting 100µl SABC stock solution with SABC Dilution Buffer at 1:100 and mixed.

Plates were washed twice with wash buffer.

100µl of standards and samples were pipetted into the wells, plates were incubated at 37°C for 90 minutes.

Plates were washed twice with wash buffer.

100µl of Biotin-labelled detection antibody solution was added to each plate and were incubated at 37°C for 60 minutes.

Plates were washed twice with wash buffer.

90µl 3,3',5,5'-Tetramethylbenzidine (TMB) solution was added to the wells and incubated at 37°C for 15 minutes.

50µl Stop Solution was added to the wells and read at 450nm immediately on the BMG Pherastar FS.

2.6.2 CH50 assay

Serum was used for the CH50 assay to avoid interference of the ion chelating effect of EDTA in complement function [139]. Complement classical pathway activity was measured using a CH50 assay (HaemoScan, cat# K002) using the manufacturer's instructions.

1ml antibody coated sheep erythrocyte concentrate was added to 5ml Dilution Buffer and mixed.

The tube was centrifuged for 5 minutes at 400 x g and the supernatant was decanted.

The pellet was resuspended with 35ml Dilution buffer.

Human serum was used in dilutions of 4, 8, 16, 32, 64 and 128 times. A serial dilution was prepared in a round bottom microplate, resulting in 50µl per well. The positive control was lysis fluid instead of plasma, the negative control was dilution buffer instead of plasma.

50 µL of test sample, reference sample and controls were pipetted in a round bottom microplate.

50 µL of the erythrocyte suspension was pipetted to each well.

The plate was covered with a sticker and incubate for 30 minutes in an incubator at 37 °C.

100 µL stop solution was pipetted to all wells.

The plate was centrifuged at 400xg for 10 minutes to pellet the unlysed cells.

100 µL of supernatant was transferred to a well of a flat bottom microplate.

Absorbance was measured at OD 415 nm.

2.6.3 Data processing

2.6.3.1 Raw data processing

All raw data was processed in RStudio (R version 4.2.1). Samples were run in duplicates and concentrations were obtained by fitting a 5PL standard curve for the Luminex assayed samples. Complement C9 was fit with a 4PL curve. For all analytes, samples were excluded where coefficients of variation (CV) were >25% between replicates. Data was log-transformed to achieve normal distribution; outliers were removed if they were 1.5 x interquartile ranges outside Q1 and Q3, respectively. Inter-plate variation was adjusted for using a linear regression model. Two of the measured analytes, C2 and C3b/iC3b, were not included in downstream analysis as <25% of samples

were quantified. For the CH50 assay a 4PL curve was fitted to the data to determine the concentration of serum at 50% haemolysis. Inter-plate variation was adjusted using sample values and an inter-plate positive control in a linear regression model.

2.6.3.2 Statistical analysis

All data processing and analyses were performed in RStudio (R version 4.0.2), scatterplots and boxplots were made using the ggplot2 package, principal component analysis (PCA) plots using the factoextra package, and the correlation matrix using the corrplot package. Normality of data was controlled using QQ-plots, density plots, histograms, Shapiro-Wilk scores, and plotted residuals. For group comparisons in the complement protein quantifications, ANOVA with subsequent Tukey's post hoc test was used. For the CH50 assay a Student's T-test was used. For correlations, Spearman's ρ was used to assess with clinical parameters. Sex-specific analyses and correction for age were performed where appropriate. Group level analyses were all age and sex adjusted using linear regression models.

3. Results

3.1 Discovery study

3.1.1 Study overview

The aim of the discovery study was to identify novel plasma biomarker candidates for Parkinson's Disease (PD) and PD related cognitive decline using untargeted mass spectrometry. The overview of the discovery study is outlined in figure 17.

Firstly, the experimental method was optimised by testing different sample processing methods. Two different plasma sample fractionation methods and two different sample depletion methods were compared to determine the method that would yield the largest number of proteins. Once the method was optimised, all the plasma samples were processed according to the new protocol and were analysed on the mass spectrometer. The data output quality was assessed and pre-processed to account for technical/methodological variables. The pre-processed data was then analysed for both PD diagnostic biomarkers, and PD cognition biomarkers. Analyses were performed to find accurate diagnostic biomarkers for PD, various covariates were assessed, and PD progression markers were investigated. Furthermore, a panel of multiple markers for PD diagnosis was constructed, and pathways implicated in PD were explored. The analyses for cognitive PD biomarkers were made both cross sectionally as well as longitudinally with survival analyses. Additionally, the analyses were performed for males and females separately, and for individual peptides from the significantly changed proteins.

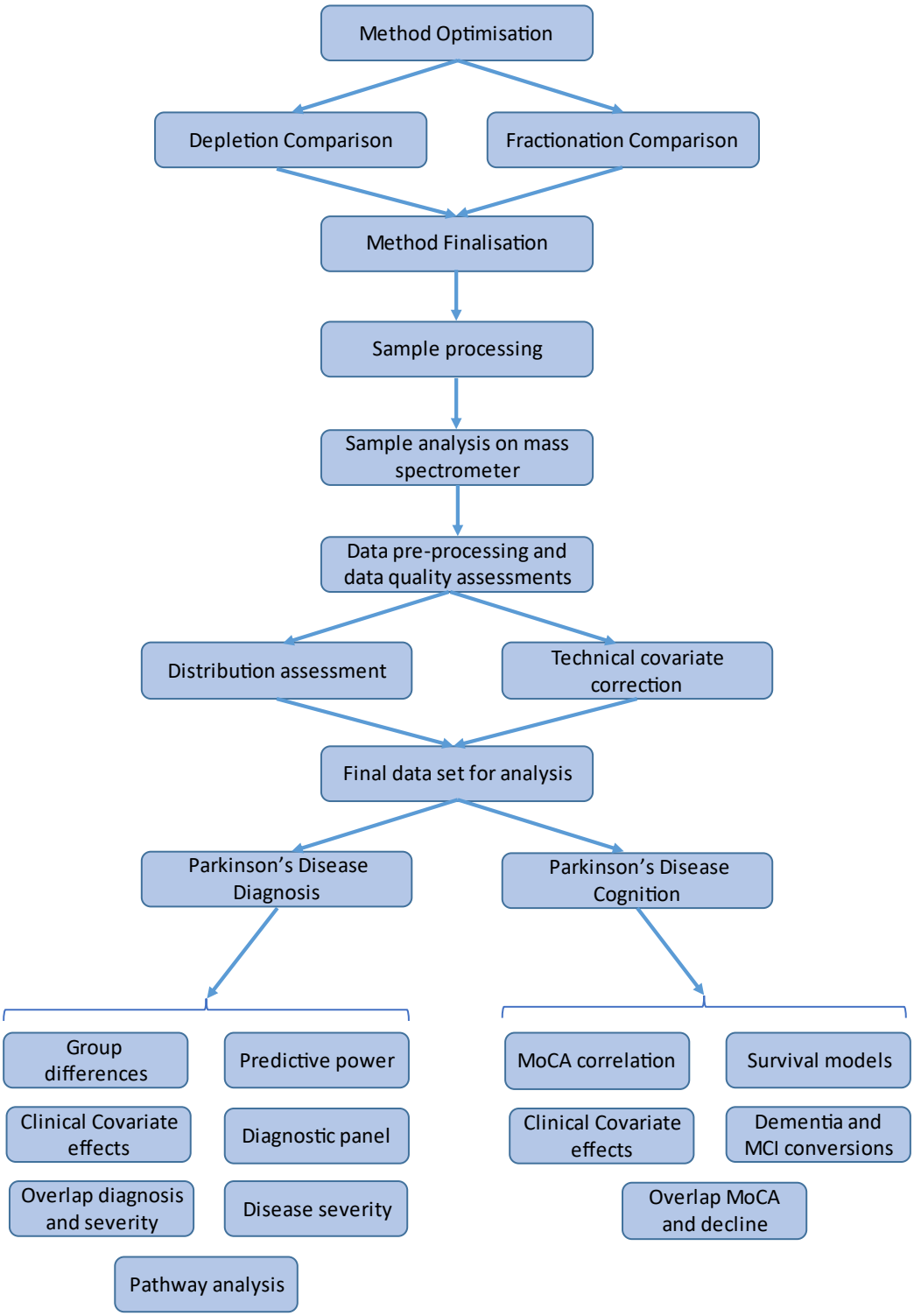

FIGURE 17 – SCHEMATIC OVERVIEW OF THE DISOVERY STUDY

3.1.2 Cohort and demographics

130 PD patients and 68 healthy controls were included in the study, demographics for the study participants are summarised in Table 4. The control subjects were sex matched, but not age matched (t-test p = 0.046. The PD cohort was clinically assessed by a movement disorder specialist for disease stage (Hoehn & Yahr), Motor score (UPDRS part III), and cognitive score (MoCA). The median Hoehn & Yahr score was 2, which means bilateral or midline involvement without impairment of balance. The median MoCA score was 25.5, which means approximately half the patients were cognitively unimpaired at time of plasma sampling, and half the patients had some level of cognitive impairment. The distribution of MoCA scores is plotted as a histogram in Figure 18. The median rate of cognitive decline was 0.5 MoCA points per year from date of diagnosis, and the median cognitive follow up after PD diagnosis was 7.9 years and is illustrated in a spaghetti plot in Figure 19.

TABLE 4 – DEMOGRAPHICS TABLE - MASS SPECTROMETRY

Demographics of Parkinson's disease patients and healthy controls. Clinical scales and other parameters include Hoehn & Yahr, Unified Parkinson's Disease Rating Scale (UPDRS) III, Levodopa Equivalent Daily Dose (LEDD), Montreal Cognitive Assessment (MoCA), Beck's Depression Inventory (BDI) II, Pittsburgh Sleep Quality Index (PSQI), Montgomery Åsberg Depression Rating Scale (MADRS), Mental Fatigue Scale (MFS), Parkinson's Disease Questionnaire (PDQ) 39, Hospital Anxiety and Depression Scale (HADS), and Non-Motor Symptom Questionnaire (NMSQ). Data presented as mean (standard deviation) or median (min – max). P-value form t-test for Age and Chi-squared for Sex.

	Parkinson's disease	*Healthy Controls*	*p-value*
Age	68.3 (9.5)	65.5 (9.0)	0.0464
Sex (F : M)	58 : 72	33 : 35	0.600
Disease duration	3.4 (0-22)	-	
Hoehn & Yahr	2 (1-4)	-	
UPDRS III	25.5 (3-80)	-	
LEDD	500 (0-2235)	-	
MoCA	25 (9-30)	-	
BDI II	10 (0-43)	-	
PSQI	7 (2-18)	-	
MADRS	7 (0-36)	-	
MFS	10.5 (0-32)	-	
PDQ 39	25.8 (0-76.6)	-	
HADS Anxiety	5 (0-19)	-	
HADS Depression	3 (0-17)	-	
NMSQ	8 (0-20)	-	

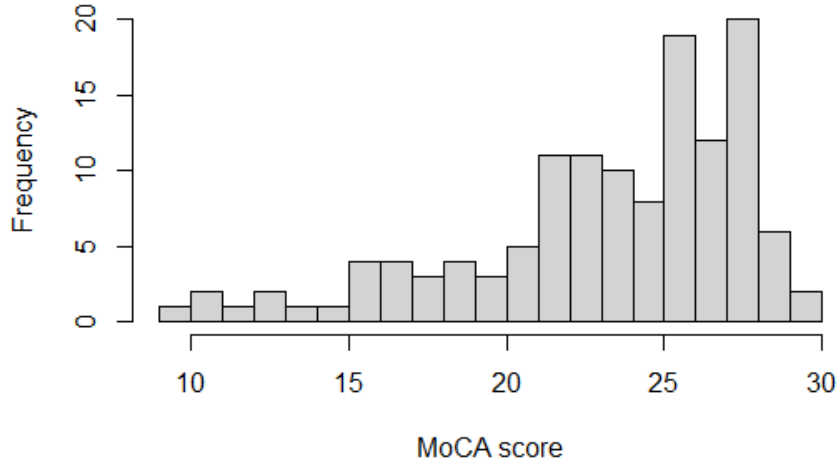

FIGURE 18 – CROSS SECTIONAL MONTREAL COGNITIVE ASSESSMENT (MOCA) SCORES FOR PARKINSON'S DISEASE (PD) PATIENTS

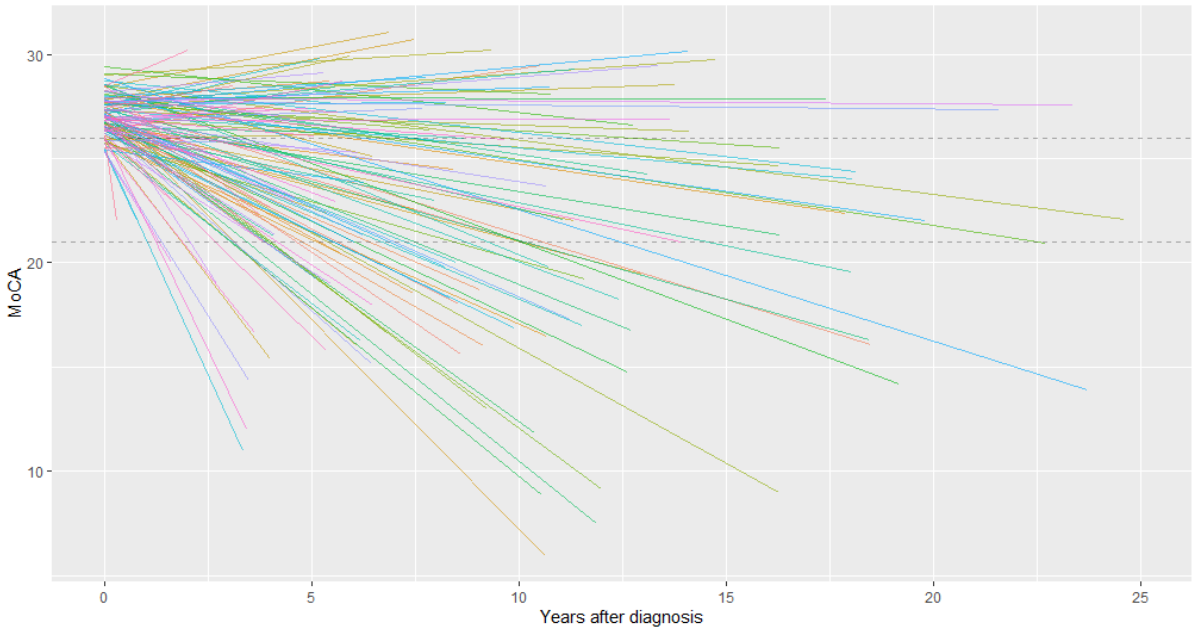

FIGURE 19 – SPAGHETTI PLOT OF LONGITUDINAL MONTREAL COGNITIVE ASSESSMENT (MOCA) SCORES FOR PARKINSON'S DISEASE (PD) PATIENTS

A PCA plot was made for the PD cohort with the clinical scores and demographic parameters (Figure 20). 41% of the variation of the data was explained by PC1, and 18% by PC2. All parameters (age, motor symptom, disease duration, non-motor symptoms) explained the variation in the same direction on PC1, which suggests some collinearity for all disease progression and severity parameters. This makes sense as PD that worsen over time, generally worsen in multiple domains. Interestingly, there was a separation in PC2 explained by motor and cognitive severity in the opposite direction from the non-motor (mainly psychometric) self-assessments. Motor symptoms, disease duration, age, LEDD, were all highly colinear. Psychometric measurements and NMS were all highly colinear.

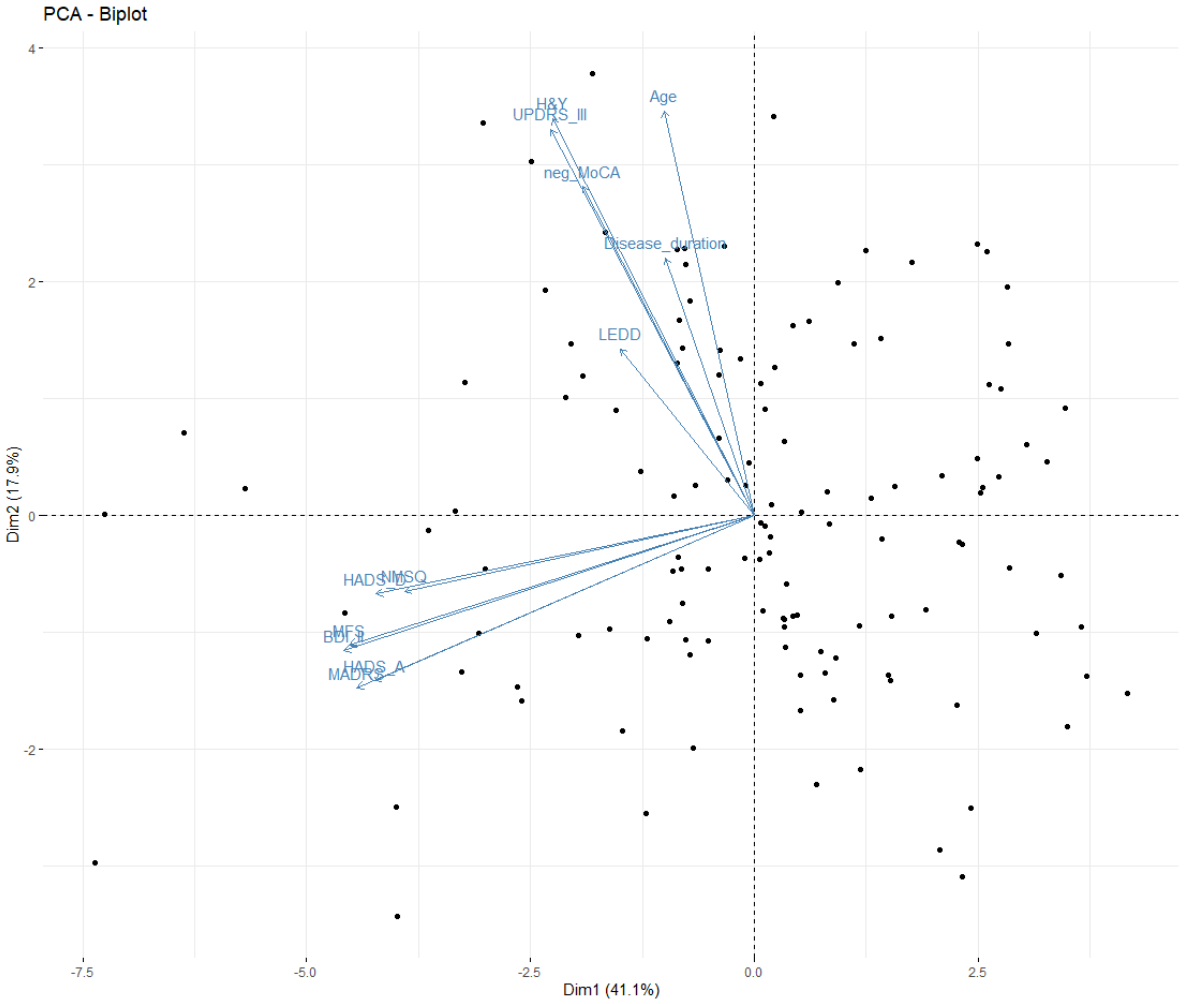

FIGURE 20 – PRINCIPAL COMPONENT ANALYSIS (PCA) OF CLINICAL AND DEMOGRAPHIC VARIABLES FOR PARKINSON'S DISEASE (PD) PATIENTS.

PCA plot with scaled and centred demographic and clinical parameters, axes show principal component 1 and principal component 2. Clinical scales and other parameters include Age, disease duration, Hoehn & Yahr, Unified Parkinson's Disease Rating Scale (UPDRS) III, Levodopa Equivalent Daily Dose (LEDD), Montreal Cognitive Assessment (MoCA), Beck's Depression Inventory (BDI) II, Pittsburgh Sleep Quality Index (PSQI), Montgomery Åsberg Depression Rating Scale (MADRS), Mental Fatigue Scale (MFS), Parkinson's Disease Questionnaire (PDQ) 39, Hospital Anxiety and Depression Scale (HADS), and Non-Motor Symptom Questionnaire (NMSQ).

3.1.3. Method optimisation

The aim of the method optimisation step was to further optimise the protocol designed by Ashton et al. [117]. Although Ashton et al. had thoroughly tested several variables in the sample preparation process, some alternative sample depletion and fractionation methods were available that had not yet been tested. Two different fractionation methods, one using reverse phase, high pH fractionation, and an isoelectric focusing (IEF) gel strip method were compared. Additionally, Top 14 depletion columns were compared with the previously used ProteoPrep (Top 2) columns for depleting the most abundant proteins and maximise detection of low abundant proteins. Finally, the effect of number of defrost cycles a sample had undergone was investigated. The protocol was setup so that all methodological variables could be tested while other variables are kept constant. 3 different plasma samples with anonymised IDs A, B, and C were used, where one aliquot of sample A had undergone 3 freeze thaw cycles, and the other aliquot 1 cycle. The variables tested for each sample are outlined in the Table 5.

TABLE 5 – EXPERIMENTAL SETUP FOR SAMPLE PROCESSING METHODS OPTIMISATION

Overview of variable comparison for method optimisation. Anonymised samples with ID A, B, and C were used. Sample A had two aliquots with differing defrost cycles. Top14 and Top2 protein depletion columns were compared. Isoelectric focusing (IEF) gel strips and spin columns (Spin) were compared as fractionation methods. The samples were labelled with 8 mass tags and run as two pooled samples.

Sample ID	Sample name	Defrost cycles	Depletion method	TMT Label	IEF or Spin	Mass spec run
14A1	A	1	Top 14	127 N 1	Spin	1
14A1	A	1	Top 14	127 C 1	IEF	1
2A1	A	1	Top 2	128 N 1	Spin	1
2A1	A	1	Top 2	128 C 1	IEF	1
2B	B	2	Top 2	129 N 1	Spin	1
2B	B	2	Top 2	129 C 1	IEF	1
2C	C	2	Top 2	130 N 1	Spin	1
2C	C	2	Top 2	126 1	IEF	1
14A2	A	3	Top 14	127 N 1	Spin	2
14A2	A	3	Top 14	127 C 1	IEF	2
2A2	A	3	Top 2	128 N 1	Spin	2
2A2	A	3	Top 2	128 C 1	IEF	2
14B	B	2	Top 14	129 N 1	Spin	2
14B	B	2	Top 14	129 C 1	IEF	2
14C	C	2	Top 14	130 N 1	Spin	2
14C	C	2	Top 14	126 1	IEF	2

The first variable investigated was the fractionation method. 302 and 329 protein IDs were identified with the IEF method at 1%FDR in the two mass spectrometry runs. The spin columns on the other hand yielded 374 and 348 protein IDs in the two runs, which is about 15% more proteins groups detected with the spin columns (Figure 21). However, the IEF method generated more than twice as many PSMs per protein on average than with the spin columns. Finally, the average protein coverage and number of unique peptides per proteins was calculated. The results were comparable between IEF and spin column fractionation, with a slightly higher coverage of the protein for the spin column method. This suggests that despite more than twice the number of mass spectra were matched with protein IDs in the IEF fractionated samples, these mass spectra were likely from a narrower set of peptides compared with the spin column fractionated samples. Taken together spin column fractionation resulted in a higher level of protein IDs and higher level of protein unique peptides, which suggests it is a superior method to IEF fractionation of plasma samples. Moreover, spin column fractionation is a far less labour-intensive method and where less sample manipulation is involved, which introduces less confounders. Hence the spin columns were ultimately selected for the discovery study.

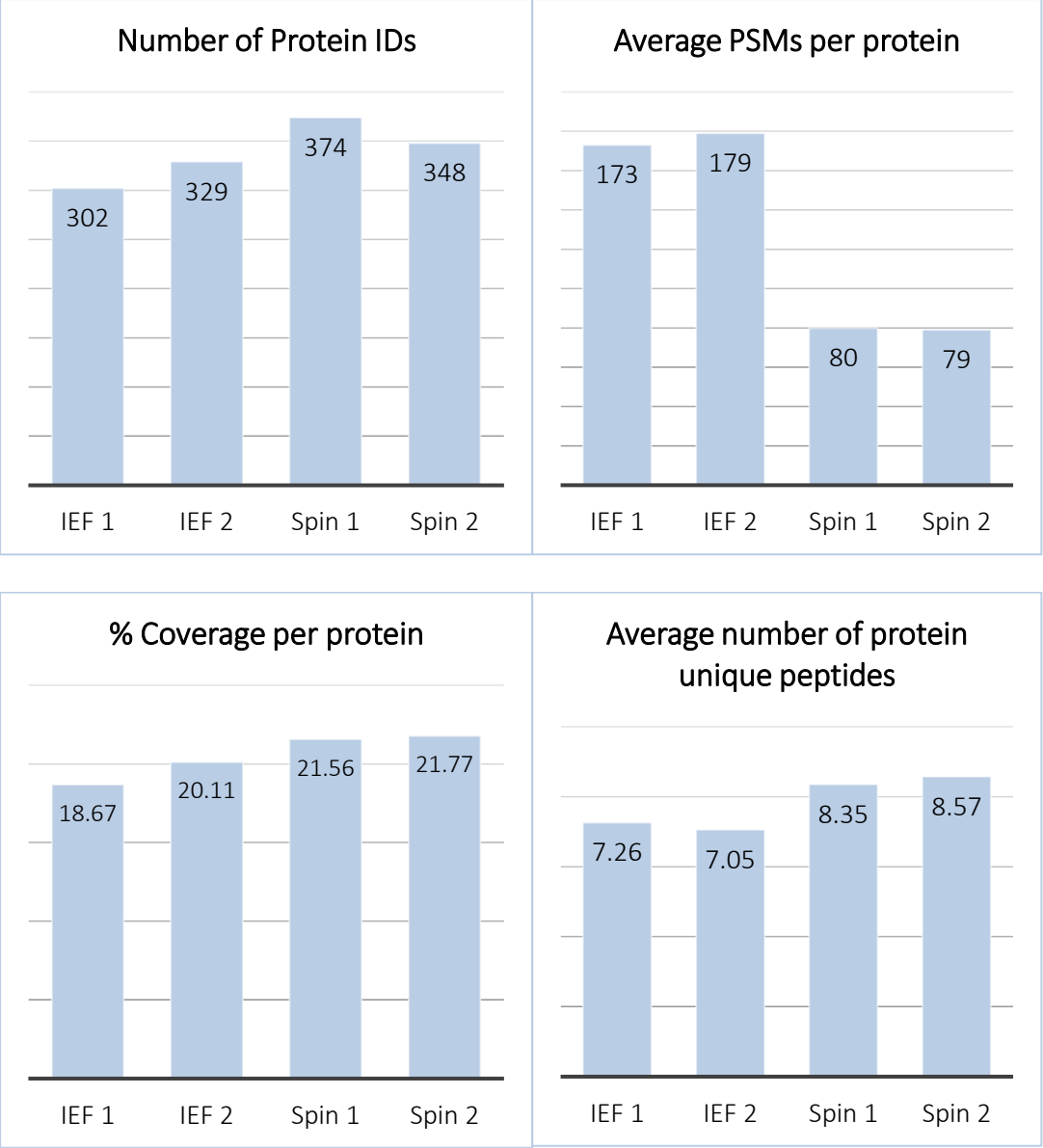

FIGURE 21 – PROTEIN DETECTION AND COVERAGE – METHODS COMPARISON

Fractionation method comparison between Isoelectric Focusing (IEF) gel strips and high pH spin columns. Number of Protein IDs, average Peptide Spectral Matches (PSMs) per protein, % protein coverage, and average number of protein unique peptides were compared.

Next, the depletion columns were compared to identify which one best depleted the high abundant plasma proteins. The two options were the Top 2 column that depleted albumin and IgG, and the Top 14 column that also deplete the 12 next most prevalent plasma proteins (IgD, IgE, IgG, IgG (light chains), IgM, Alpha-1-acid glycoprotein, Alpha-1-antitrypsin, Alpha-2-macroglobulin, Apolipoprotein A1, Fibrinogen, Haptoglobin, and Transferrin). The total TMT counts were summed for all spin column

fractions for the 4 samples A1, A2, B and C. Ratios between Top 2 and Top 14 TMT counts for all proteins was calculated by dividing the TMT value for the Top 2 depleted sample with the Top 14 depleted sample. The ratios were plotted for each protein that was listed as targets for the Top 14 columns in Figure 22. The immunoglobulin subunits were all summed up. TMT values were also summed for any other protein that had multiple subunits quantified. As seen in Figure 22 the protein levels based on TMT scores were approximately half for most proteins using the Top 14 spin columns. If the assumption is made that TMT levels are linearly associated with protein abundance, this suggests approximately half the targeted proteins were depleted with the Top 14 spin columns. Interestingly, the average albumin concentration was around 9-fold higher in the Top 14 depleted samples, which suggests they had much poorer albumin depletion compared with the Top 2 columns. Given that depletion of albumin, the by far most prevalent plasma protein, was inefficiently depleted with the Top 14 columns, and the TMT counts suggest only about half the abundance of the remaining proteins was depleted, suggests Top 2 columns are the better choice for the discovery study.

Finally, it was investigated whether freeze thaw cycles had any major effect on the proteins quantified. The median coefficient of variation (CV) for the quantified proteins in the spin column fractionated samples was 15%, which suggests that the there is only minor alterations in protein levels between 1 and 3 freeze thaw cycles. Interestingly, it was found that for the same IEF fractionated samples (i.e. sample A with 1 versus 3 freeze thaw cycles) the median CV was 34% which suggests larger methodological variation using IEF fractionation, which further speaks in favour of the spin column fractionation method.

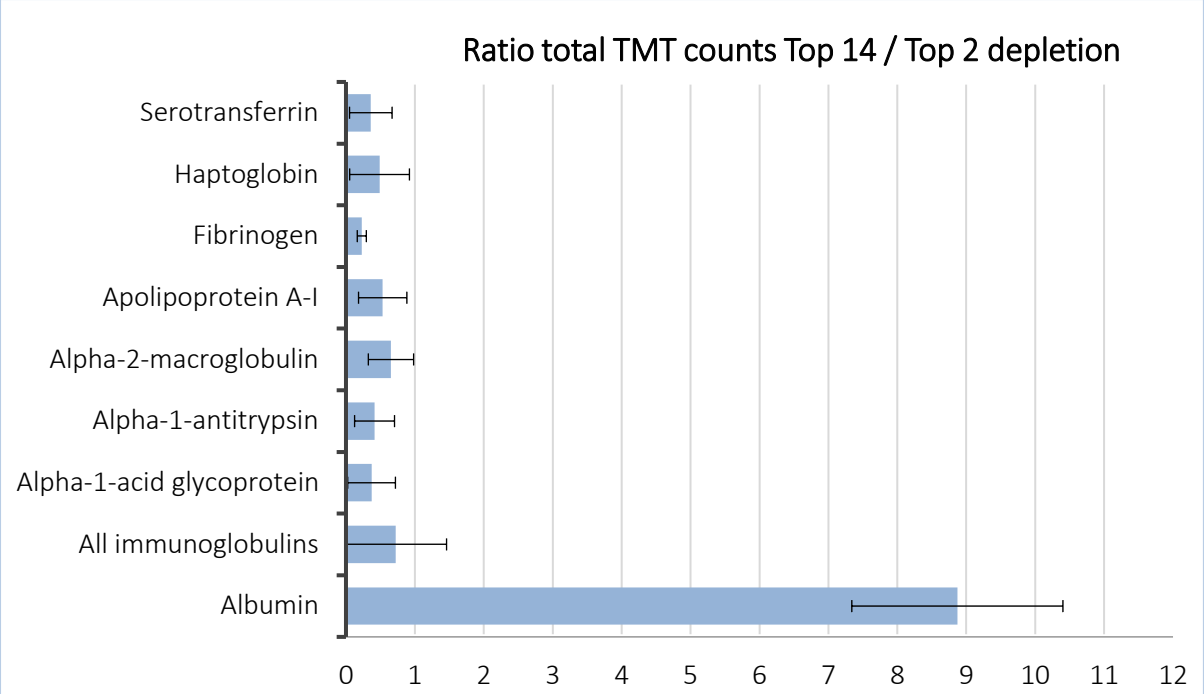

FIGURE 22 – DEPLETION COLUMN COMPARISON

Figure showing comparison between Top 14 and Top 2 (ProteoPrep) depletion columns as the protein abundance ratio between sample for High pH Reversed-phase Peptide Fractionation column samples. Bars represent mean and standard deviation for the two replicates.

In conclusion the Top 2 depletion method (ProteoPrep) was chosen owing to more efficient albumin depletion, and the spin columns were chosen for the fractionation owing to higher number of protein IDs and more protein unique peptides detected. The final sample processing protocol for the discovery study is summarised in Figure 23.

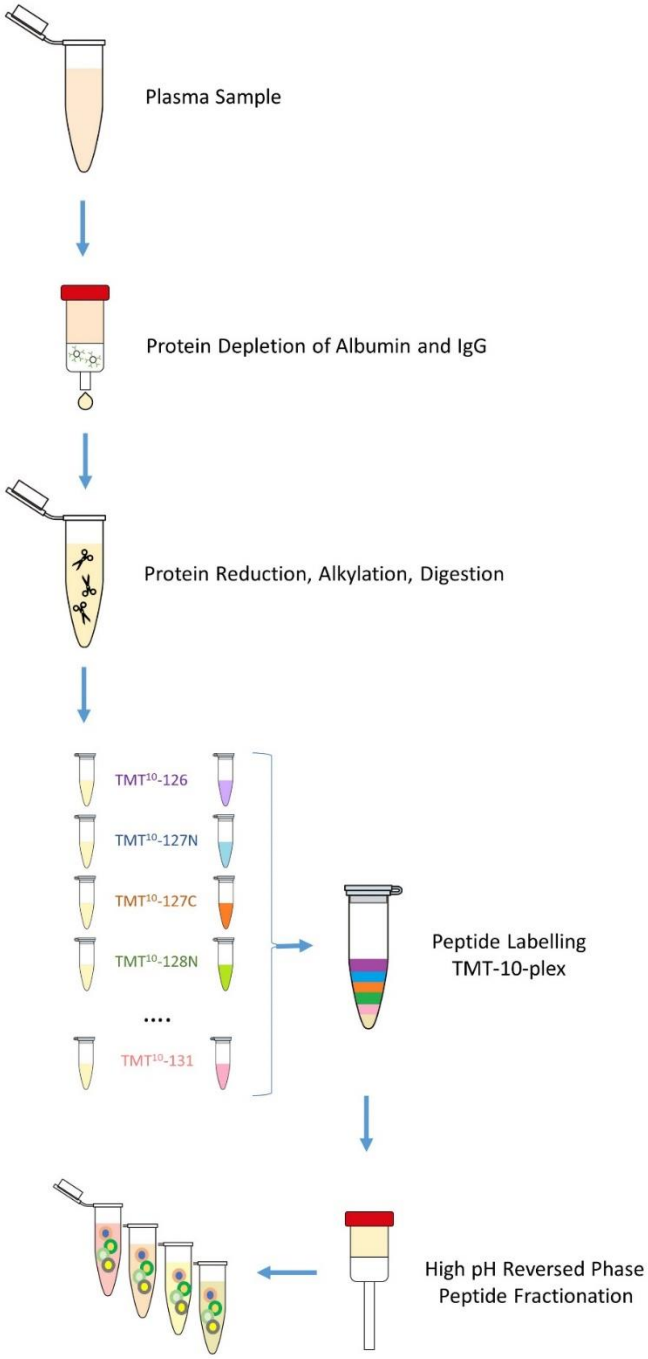

FIGURE 23 – SAMPLE PREPARATION WORKFLOW – FINAL PROTOCOL

Schematic overview of the final workflow used for processing samples in the discovery study. In brief, plasma samples were depleted from albumin and IgG with antibody-based columns. They were then chemically reduced and alkylated, and digested with trypsin. Samples were finally labelled with TMT mass tags and fractionated with spin columns by hydrophobicity.

3.1.4 Discovery study – untargeted proteomic on the study cohort

3.1.4.1 Raw data Pre-processing

The 198 plasma samples from 130 PD patients, and 68 healthy controls were processed according to the optimised sample preparation protocol (Figure 23). The samples were analysed on the mass spectrometer, and the raw data output was processed for peptide identification and quantification. The mass spectra were matched against Mascot and Sequest proteome databases, and the TMT reporter ions were quantified.

20,072 protein groups were identified at a 5% FDR. Their abundances were determined with the TMT counts. Prior to using protein quantifications in the analysis with the clinical data, the data quality needed to be assessed and pre-processed appropriately.

First, the distribution of total protein abundances for each sample was visualised. As seen in in Figure 24, the distributions of the protein abundances were heavily positively skewed. This meant the number of proteins at lower abundances were much higher than the proteins at higher abundances. This was expected and in line with what is known about protein distribution in plasma.

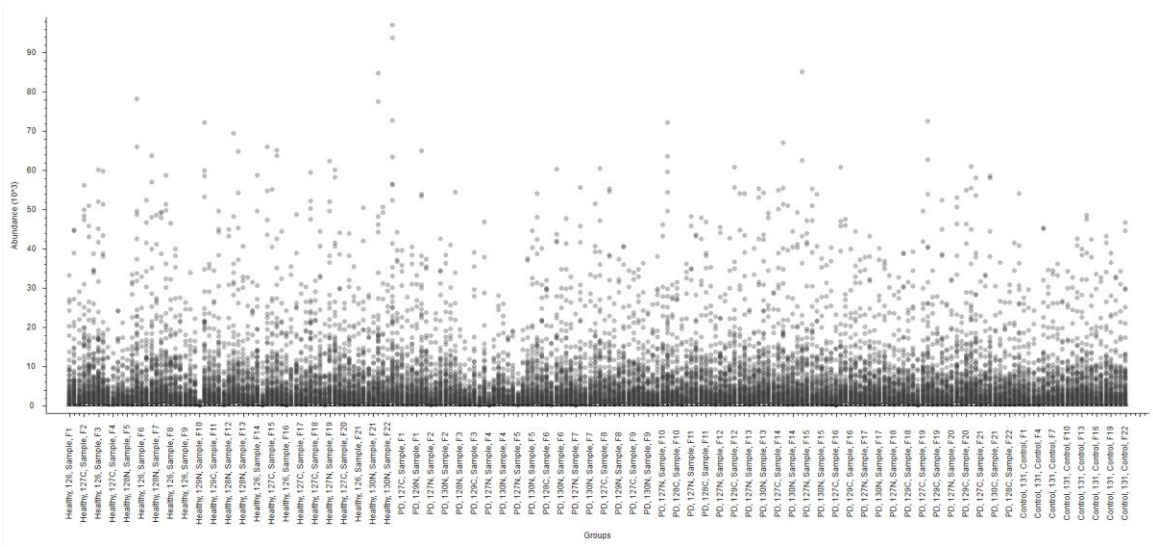

FIGURE 24 – TOTAL PROTEIN ABUNDANCE DISTRIBUTION ACROSS ALL MASS SPECTROMETRY SAMPLES

Boxplots showing total (arithmetic) protein abundance (y-axis) for each sample (x-axis).

Abundances were subsequently log10 transformed to obtain closer to a normal distribution of protein abundances. It was observed that the total protein abundance was relatively even across samples, with the extremes approximately having medians one order of magnitude apart. The quality control samples (blue) showed little variation across replicates. This suggests that there was relatively little variation in total proteins between sample runs owing to technical variation, and the inter sample total abundance variation in Figure 25 is mostly due to intrinsic sample properties.

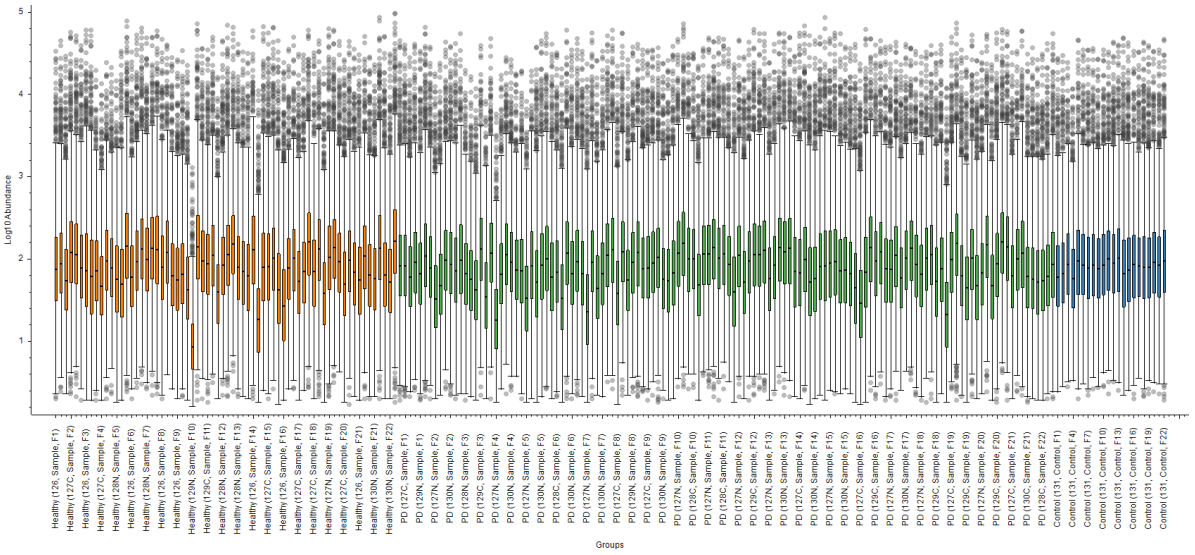

FIGURE 25 – TOTAL PROTEIN ABUNDANCE DISTRIBUTION ACROSS ALL MASS SPECTROMETRY SAMPLES - LOG Y-AXIS

Boxplots showing total (logarithmic) protein abundance (y-axis) for each sample (x-axis). Orange = healthy controls, Green = Parkinson's disease, blue = Quality control sample

As the relative abundance of an individual protein is dependent on the total amount of protein quantified, the total protein abundance was subsequently normalised across all samples in Proteome Discoverer (ver. 2.3; Thermo Scientific).

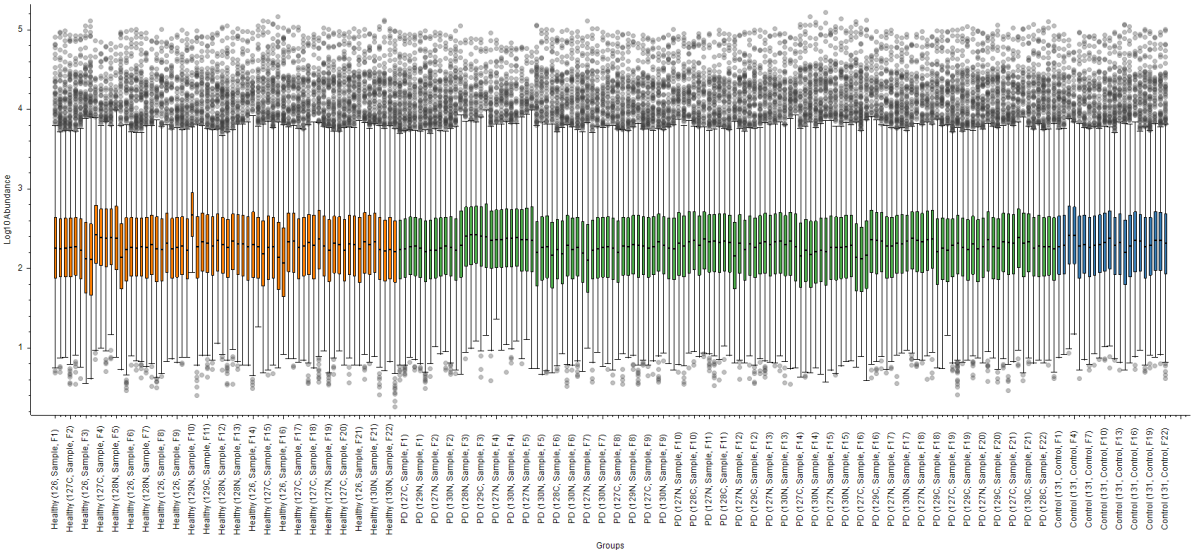

FIGURE 26 – TOTAL PROTEIN ABUNDANCE DISTRIBUTION ACROSS ALL MASS SPECTROMETRY SAMPLES - LOG Y-AXIS AND NORMALISED ABUNDANCES

Boxplots showing normalised (logarithmic) protein abundance (y-axis) for each sample (x-axis). Orange = healthy controls, Green = Parkinson's disease, blue = Quality control sample

Except for few outliers, this normalisation step generated even total abundances across all 220 sample IDs (Figure 26). It was finally investigated that there were no apparent discrepancies post normalisation between the different mass tags or between 10-plex runs. As seen in Figure 27 there did not seem to be any noticeable technical batch difference with regards to TMT tag nor 10-plex sample run that affected the abundance distribution.

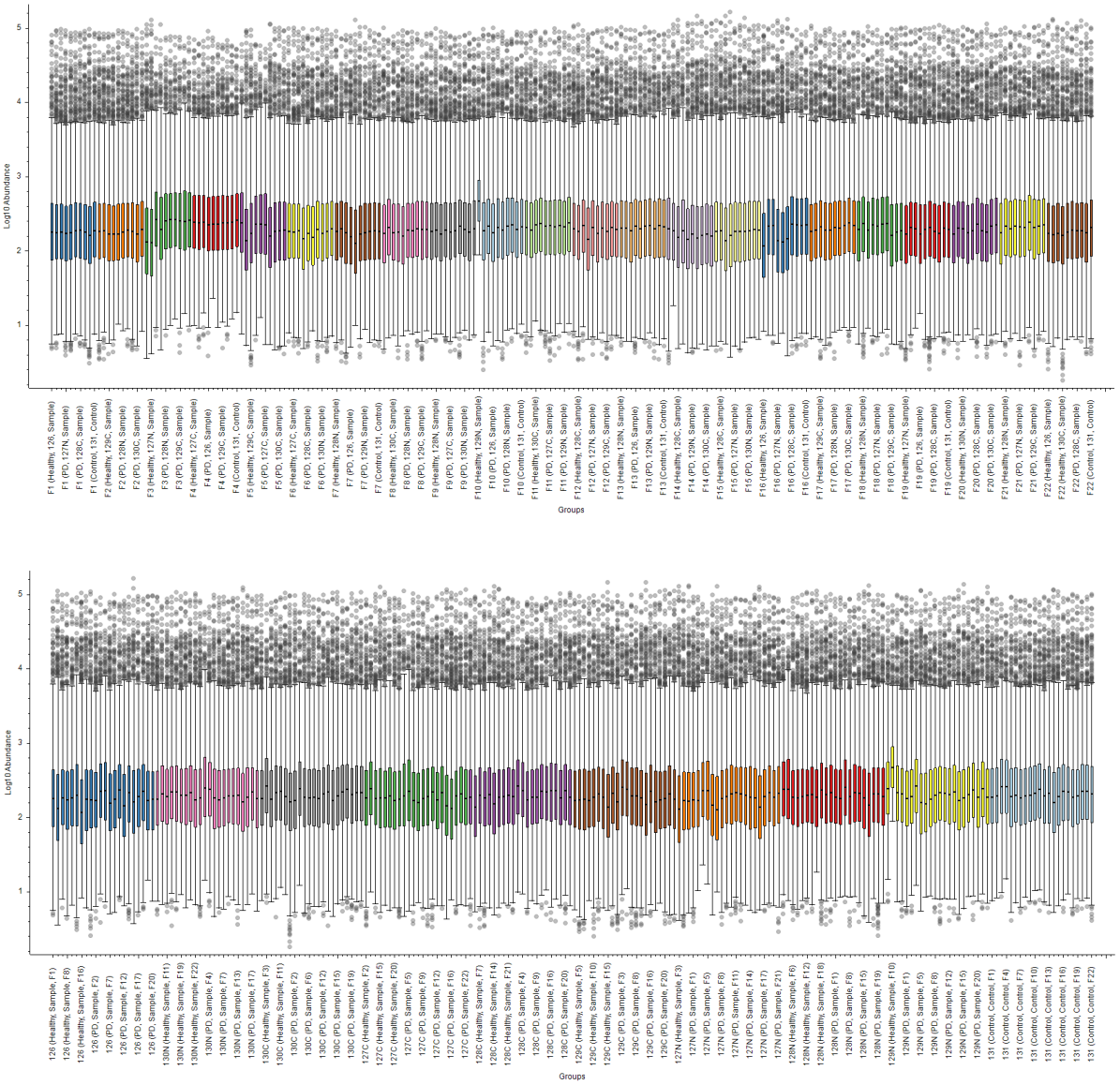

FIGURE 27 – TOTAL PROTEIN ABUNDANCE DISTRIBUTION ACROSS ALL MASS SPECTROMETRY SAMPLES - LOG Y-AXIS AND NORMALISED ABUNDANCES.

Top – coloured by sample group. Bottom – coloured by TMT tag

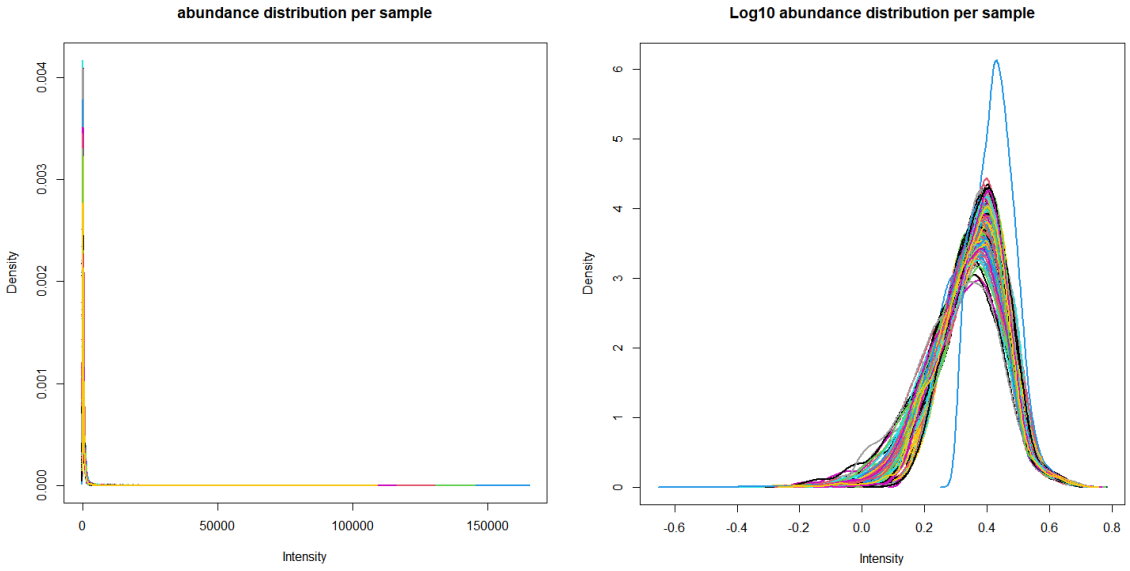

FIGURE 28 – TOTAL PROTEIN ABUNDANCE DISTRIBUTION ACROSS ALL MASS SPECTROMETRY SAMPLES – AS DENSITY PLOTS.

Left – arithmetic density plot. Right – log10 transformed density plots.

Similarly, plotting the sample abundances as density plots (Figure 28) revealed most samples had approximately normal distribution when log10 transformed. Again, one outlier appeared, whereas the rest of the samples displayed a normal distribution.

Finally, the distributions of each of the individual proteins were visualised where the protein was quantified in at least 50% of study individuals (n=2263) as density plots. It was observed that nearly all proteins displayed a normal distribution. Figure 29A shows protein distributions across approximately 4 orders of magnitude as quantified by the mass spectrometer, and an increase in number of proteins toward the lower detection level.

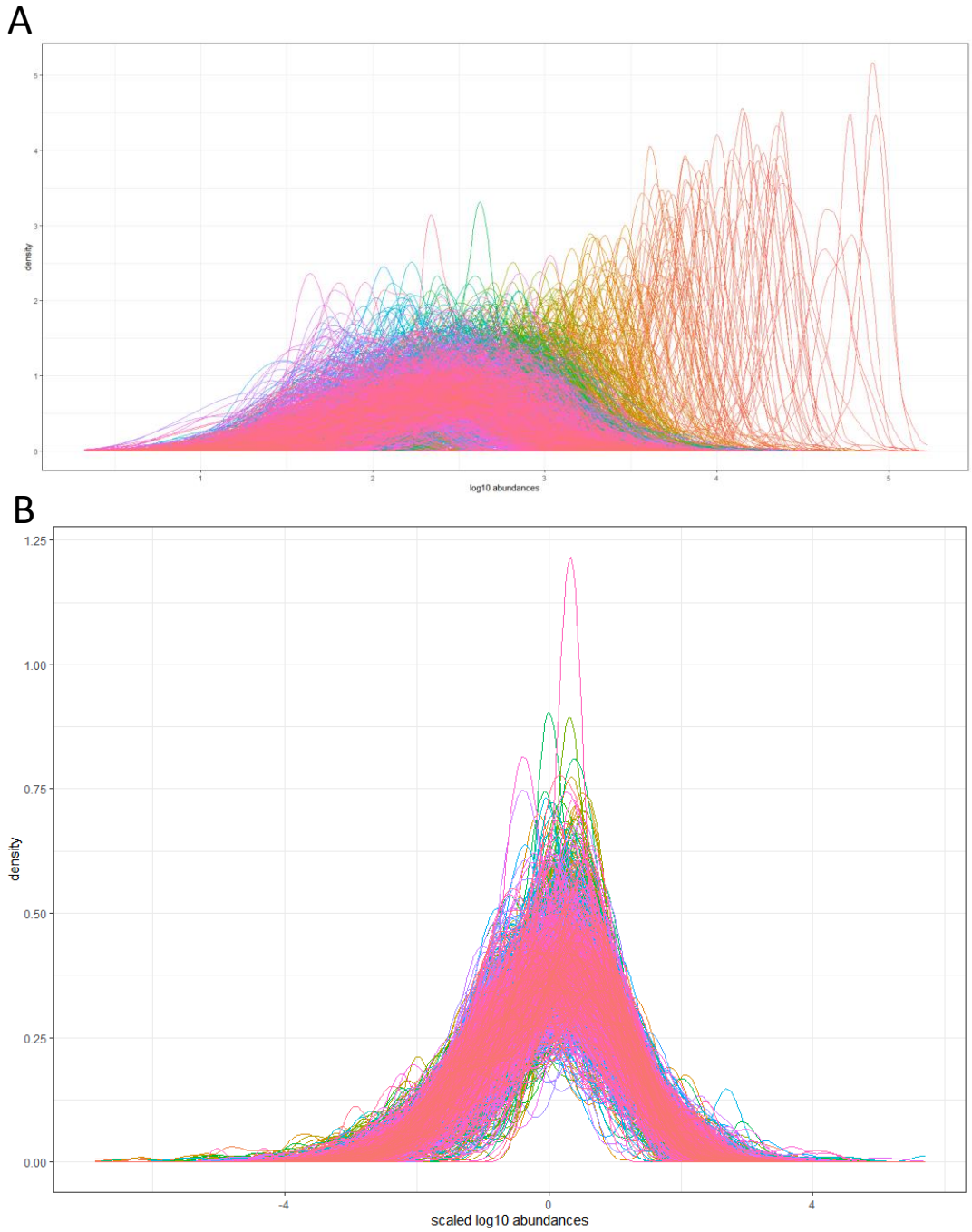

FIGURE 29 – DENSITY PLOTS FOR EACH PROTEIN THAT WAS QUANTIFIED IN >50% OF SAMPLES

A) *Density plots of protein distributions for all quantified protein IDs in >50% of samples. B) Density plots of protein distributions, scaled and centred for all proteins.*

The distributions for each protein were scaled and centred, and again appeared comparable, irrespective of their abundance levels (Figure 29B). However, as the plot displays 2263 proteins, many of which were overlapping, visual inspection of each protein was not ideal. A Shapiro wilk analysis of each protein distribution was performed to get an approximate overview of which proteins were

normally distributed. A histogram (Figure 30) revealed approximately 15% of samples had a p-value < 0.05, and 85% of proteins would be considered not deviating significantly from normally distributed across all samples.

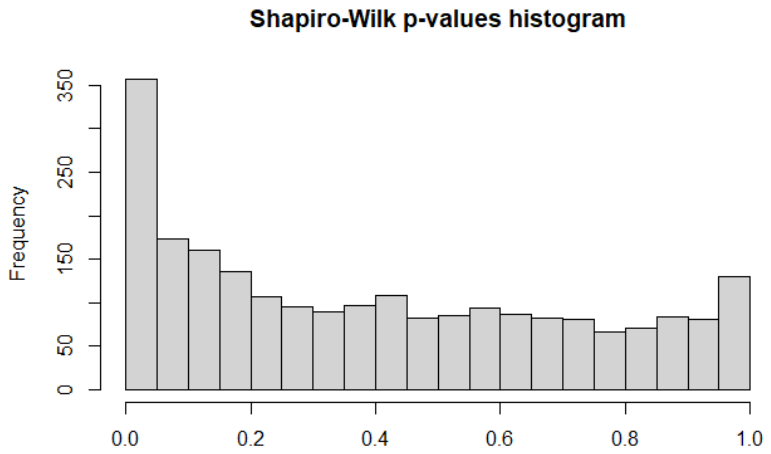

FIGURE 30 – P-VALUE HISTOGRAM FOR SHAPIRO WILK P-VALUES – DISTRIBUTION OF EACH PROTEIN

Next, the distributions of protein levels were centred and plotted as horizontal boxplots for all study samples and quality controls (QC) separately (Figure 31A). This was done in order to assess the variation in the data relative to the QC sample variation, and if further correction was needed. The 198 individual plasma samples were run in as 22 10-plex samples, where plasma samples were from 9 individuals was combined with one inter-sample quality control (QC). The QC sample was a plasma mix made of equal amount of each of the 198 samples and used as an inter-run control. It was observed that although variation of all plasma samples was greater than that of QC samples, the QC samples still displayed a large variation and needed to be corrected for. The coefficient of variation (CV) for all QC samples were therefore plotted as a histogram in Figure 31B, and it was noted the average coefficient of variation for a protein in the control sample was 15%, which is relatively large. Moreover, the median range of a protein quantity across all QCs span 2 orders of magnitude. This warranted additional correction for QC protein levels which was performed.

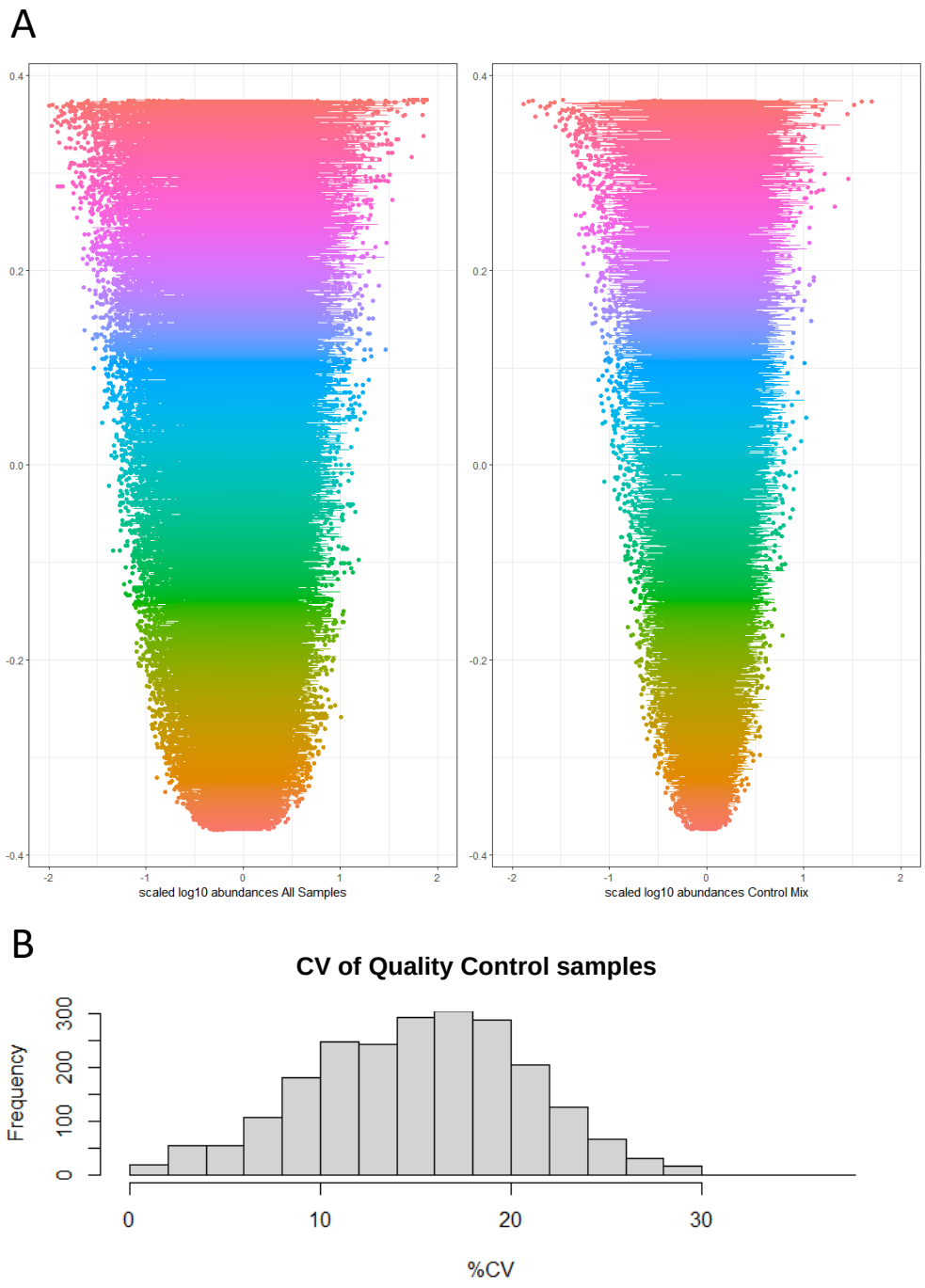

FIGURE 31 – PROTEIN DISTRIBUTIONS OF SAMPLES

A) *Boxplots of all proteins quantified, ranked by range on y-axis and centred log10 abundance on x-axis. Left – protein distributions in all samples. Right – protein distributions in quality control samples. B) Histogram of percentage coefficient of variation of quality control samples.*

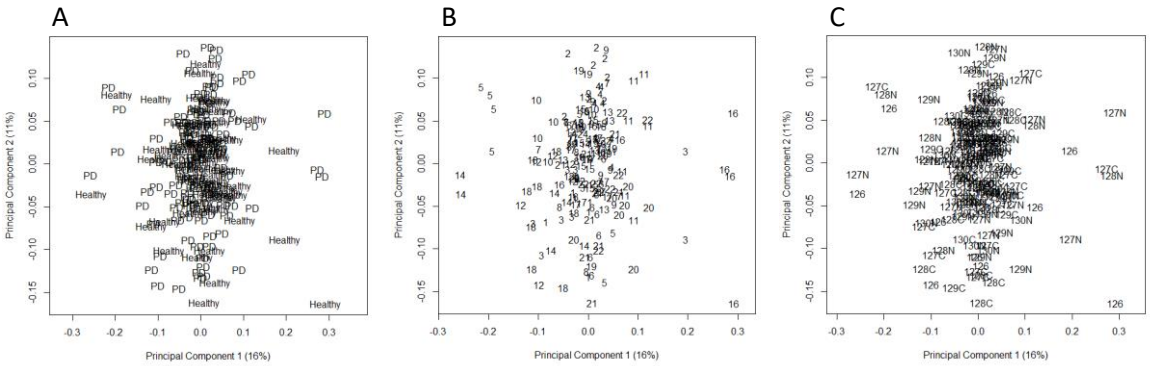

FIGURE 32 – MODIFIED PRINCIPAL COMPONENT ANALYSIS (PCA) PLOTS OF DIAGNOSIS AND TECHNICAL COVARIATES

A) *PCA of PD and healthy controls. B) PCA of 10-plex sample run (1-22). C) PCA of TMT tags used.*

Subsequently, it was explored whether any batch effects or group effects were present in the dataset. Modified PCA plots using the limma multidimensional scaling plots were used to explore group effects by 10-plex sample and TMT-label. A technical group effect was observed in the samples (Figure 32B), and this was corrected for using the limma "removeBatchEffect" formula.

Finally, a quartile normalisation was attempted to further normalise distributions. However, this altered the quantifications and results too much, and would not reflect the real-life protein distribution variations seen in human plasma and was not implemented in the pre-processing.

At this stage, the data was considered sufficiently pre-processed to be analysed with the clinical data. Hence is it was investigated whether any of these proteins would be associated with PD diagnosis and cognitive decline.

For all the analysis, all proteins were eliminated where they were quantified in less than 50% of the samples. This was done as the proteins under this threshold were more likely to have too low confidence in the proteome database match. These proteins were generally lower abundant and had fewer unique peptides matched with a lower confidence. This reduced the total number of proteins to 2260, which was about 11% of all proteins identified, or about 19% of all quantified proteins (Figure 33).

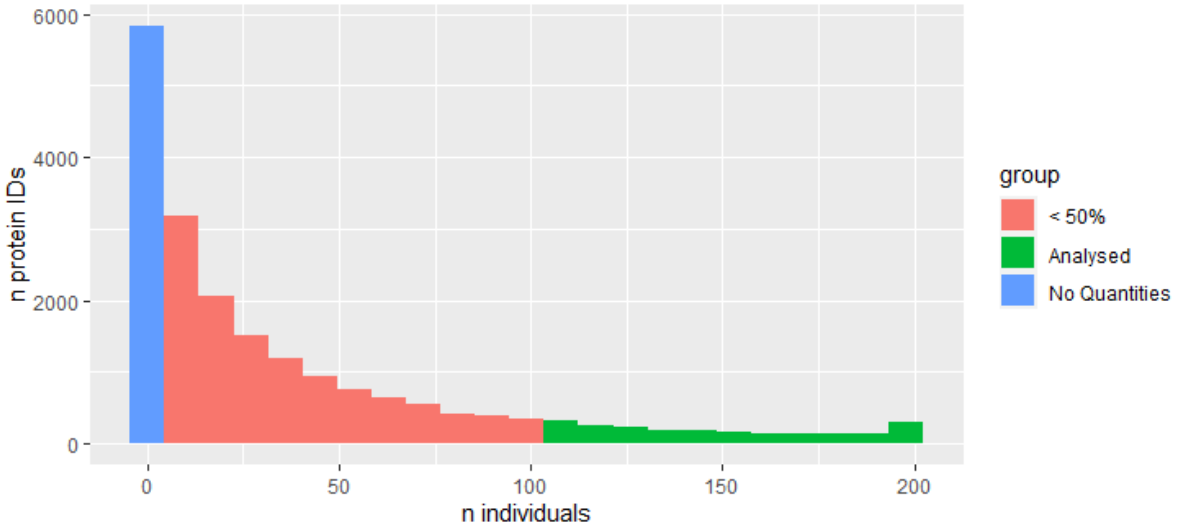

FIGURE 33 – HISTOGRAM OF NUMBER OF PROTEIN IDS THAT WERE QUANTIFIED IN DIFFERENT NUMBER OF STUDY INDIVIDUALS.

Histogram showing the number individuals where a certain number of Protein IDs were quantified in the discovery study. Total 20,072 proteins were detected, out of which 5,835 were not quantified (blue), 11,977 were quantified in <50% of individuals (red), and 2260 were quantified in more than half the individuals (green).

To further quality control our dataset, quantifications of known established blood biomarkers were used as positive controls. As there are no established blood biomarkers for PD that are relatively high abundant in plasma, proteins that were well known to alter with demographic variables were analysed instead. Von Willebrand Factor (VWF) [187], Cystatin C (CST3) [188] and Fibrinogen (FGA) [189] are all well documented to increase with age. This was in accordance with our findings, as all three of VWF ($r = 0.27$, $p = 0.00011$), CST3 ($r = 0.28$, $p = 0.00011$), and FGA ($r = 0.29$, $p = 0.000039$) were some of the most age associated proteins (Figure 34).

Next, it was investigated which proteins were most strongly associated with sex. Again 3 of the top sex associated proteins from our dataset were well documented to be sex-dependent in the literature. These were Pregnancy Zone Protein ($t = 10$, $p < 1.6 \times 10^{-20}$) [190], Ceruloplasmin ($t = 6.1$, $p = 6.0 \times 10^{-9}$) [191], and Apolipoprotein E ($t = 4.8$, $p = 3.1 \times 10^{-6}$) [192], all of which are known to be elevated in female plasma (Figure 35).

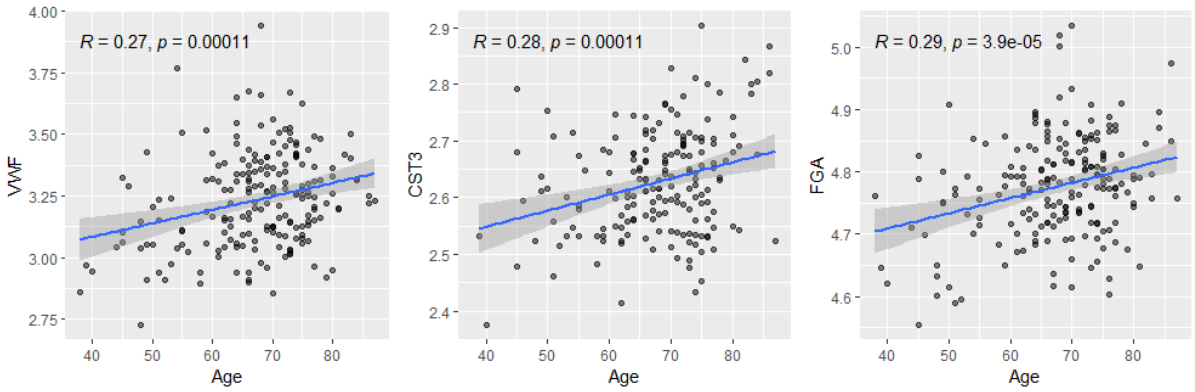

FIGURE 34 – CORRELATION BETWEEN AGE AND COMMONLY AGE ASSOCIATED PROTEIN LEVELS

Plasma levels of Von Willebrand Factor (VWF), Cystatin C (CST3), and Fibrinogen Alpha Chain (FGA), and their correlation with Age in study individuals. Pearson r and p-value shown.

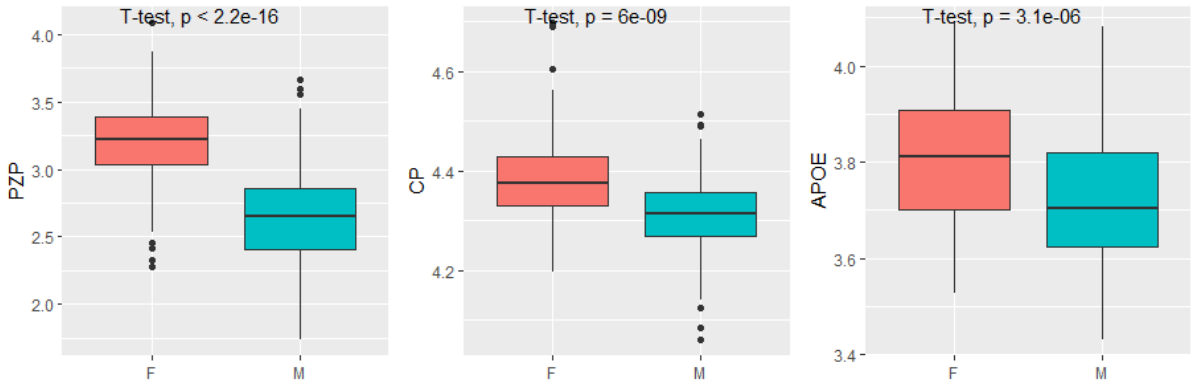

FIGURE 35 – PLASMA PROTEOMIC SEX DIFFERENCES

Plasma levels of Pregnancy Zone Protein (PZP), Ceruloplasmin (CP), and Apolipoprotein E (APOE), and their Sex differences in study individuals. M = Male, F = Female, p-value for t-test shown.

To summarise, the raw data generated from the untargeted proteomic experiment was pre-processed and quality controlled, before continuing with the analysis. The raw data was log transformed and adjusted for total protein abundance. Distributions across all samples, technical replicates, TMT-tags, and study groups were assessed. Protein distributions were visualised, and quantities corrected for using an inter-run QC sample. Minor remaining technical group effects were corrected for, and protein IDs that were quantified in less than half the study participants were eliminated. As positive controls for our study population it was inspected whether well documented plasma proteins were affected with age and sex as described in the literature.

3.1.4.2 PD diagnostic biomarkers

3.1.4.2.1 DIMENSIONALITY REDUCTION

Once the data was pre-processed, it was first explored whether the human plasma proteome could differentiate PD from healthy control individuals. To get an overview of the whole proteome, the data was transformed onto two dimensions both with a PCA (linear dimensionality reduction) (Figure 36) as well as t-SNE (non-linear dimensionality reduction) (Figure 37). No apparent separation of the data was visualised with either of the methods. A slight shift of the data along PC1 (which accounted for 16.7% of the data) was observed, but the confidence intervals were largely overlapping, and no apparent clustering was observed in the t-SNE plot. This was not surprising as the human plasma reflect all processes in the body and would not necessarily be shifted by a neurological disorder.

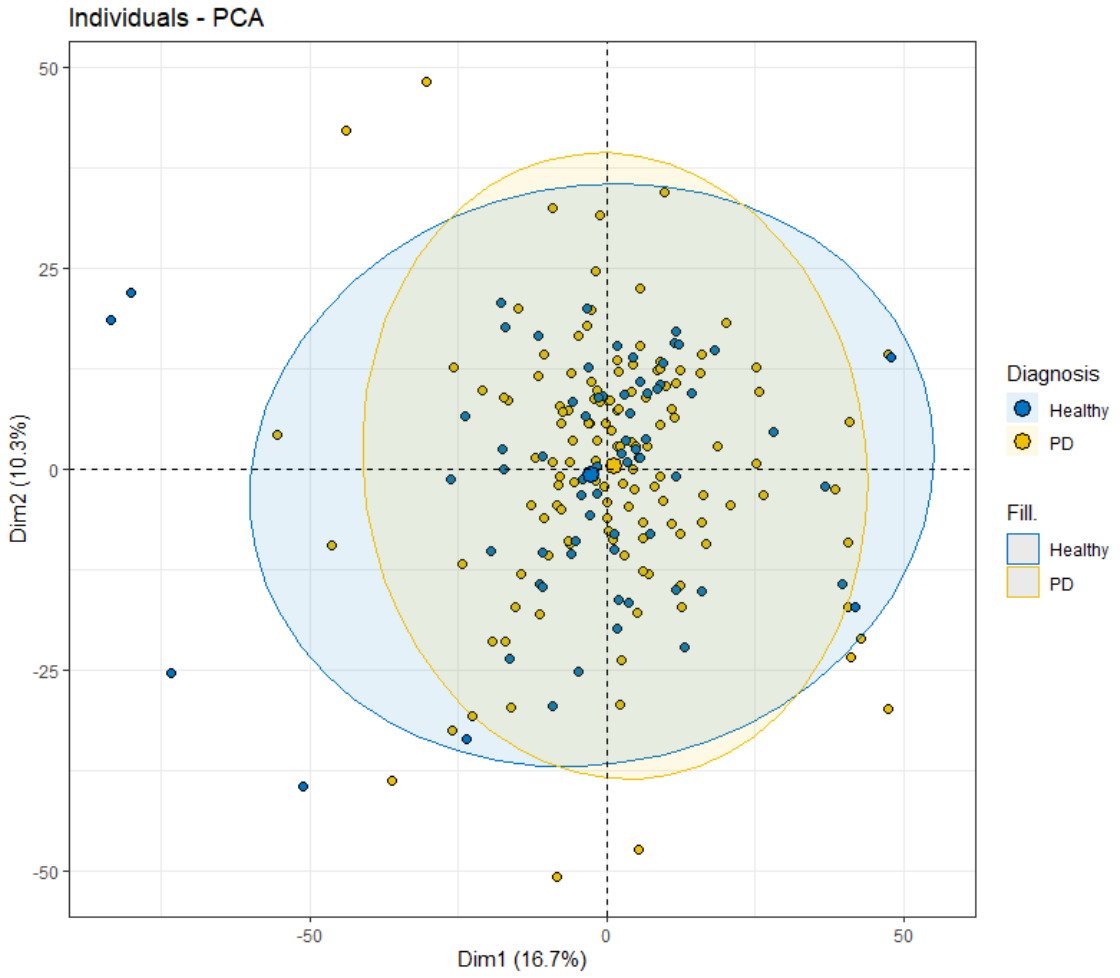

FIGURE 36 – PRINCIPAL COMPONENT ANALYSIS (PCA) PLOT OF PROTEOMIC DATA

PCA plot of mass spectrometric data, coloured by diagnosis Parkinson's disease (PD) or healthy controls. Centre point enlarged; ellipse shows 95% confidence interval. Principal components 1 and 2 shown. Variables are centred and scaled.

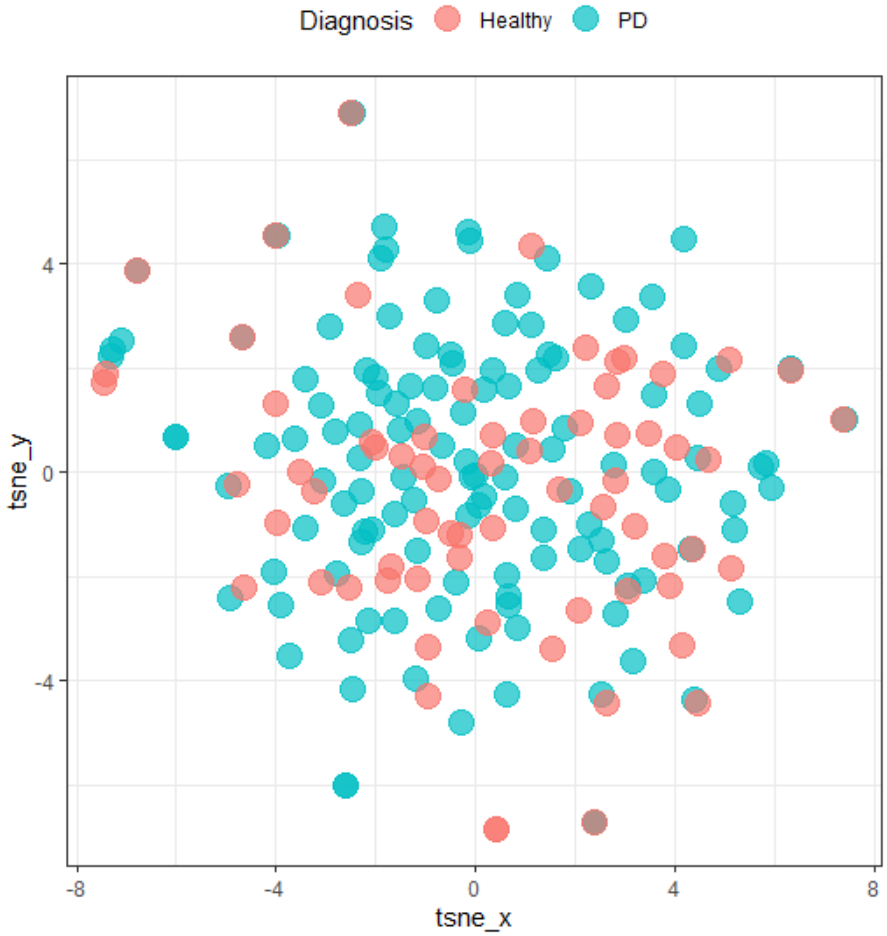

FIGURE 37 – T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (TSNE) OF PROTEOMIC DATA.

t-SNE plot of mass spectrometric data, coloured by diagnosis Parkinson's disease (PD) or healthy controls. Variables are centred and scaled.

3.1.4.2.2 PD VERSUS CONTROLS

Next, differential protein expression was analysed using the limma package in R. A robust linear model was fitted without an intercept, using Diagnosis (PD or healthy controls), age and sex as explanatory variables, was fitted against all protein expression levels. The contrast matrix of interest (PD versus healthy control) was applied to the data, and statistical output were calculated with empirical Bayes moderation.

A p-value histogram was plotted to visualise whether there were true differentially expressed proteins in the dataset (Figure 38). The higher-than-average bar representing p-values between 0-0.05 on the x-axis suggest that there is a true difference in protein levels for several (approximately 50-100) proteins between PD and controls, and that the null hypothesis should be rejected for this number of protein IDs.

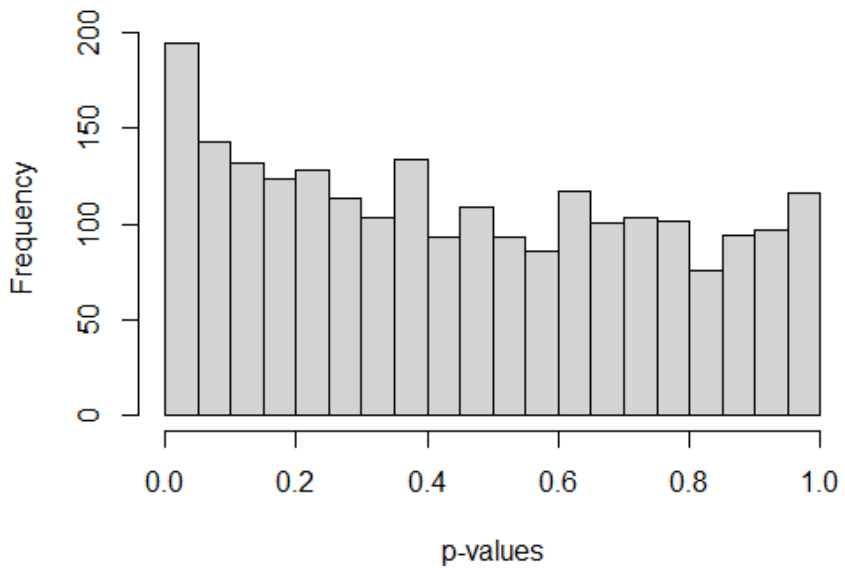

FIGURE 38 – P-VALUE HISTOGRAM FOR GROUPWISE COMPARISON OF PROTEINS BETWEEN PARKINSON'S DISEASE AND HEALTHY CONTROLS.

Results from the limma robust linear models were visualised with a volcano plots and statistics were summarised in tables. The first a model was fitted with only PD and controls, without accounting for any covariates. The results are shown in Figure 39 and Table 6. Secondly, a model was fit with age and sex as covariates, and results are displayed in Figure 40 and Table 7. It was observed that the top hits were similar between the two models. This was likely due to that the PD and Control groups were not deviating too much in terms of Age and Sex distribution. The results tables suggest very few proteins remain statistically significant after correction for age, sex and multiple comparisons using Benjamini-Hochberg. However, this correction is arguably too stringent when identifying candidate biomarkers. Since the candidate markers will be verified either way, it is not necessary to eliminate all false positives at this stage. The correction eliminates false positives to find the statistically true positive differences, but results at the same time in many false negatives. Moreover, the p-value histogram suggests that at least 50 proteins are likely differentially expressed in the dataset between PD and controls. Therefore, when implementing a discovery-verification-validation pipeline one should rather select the most significantly differentially expressed proteins as biomarker candidates that need to be validated.

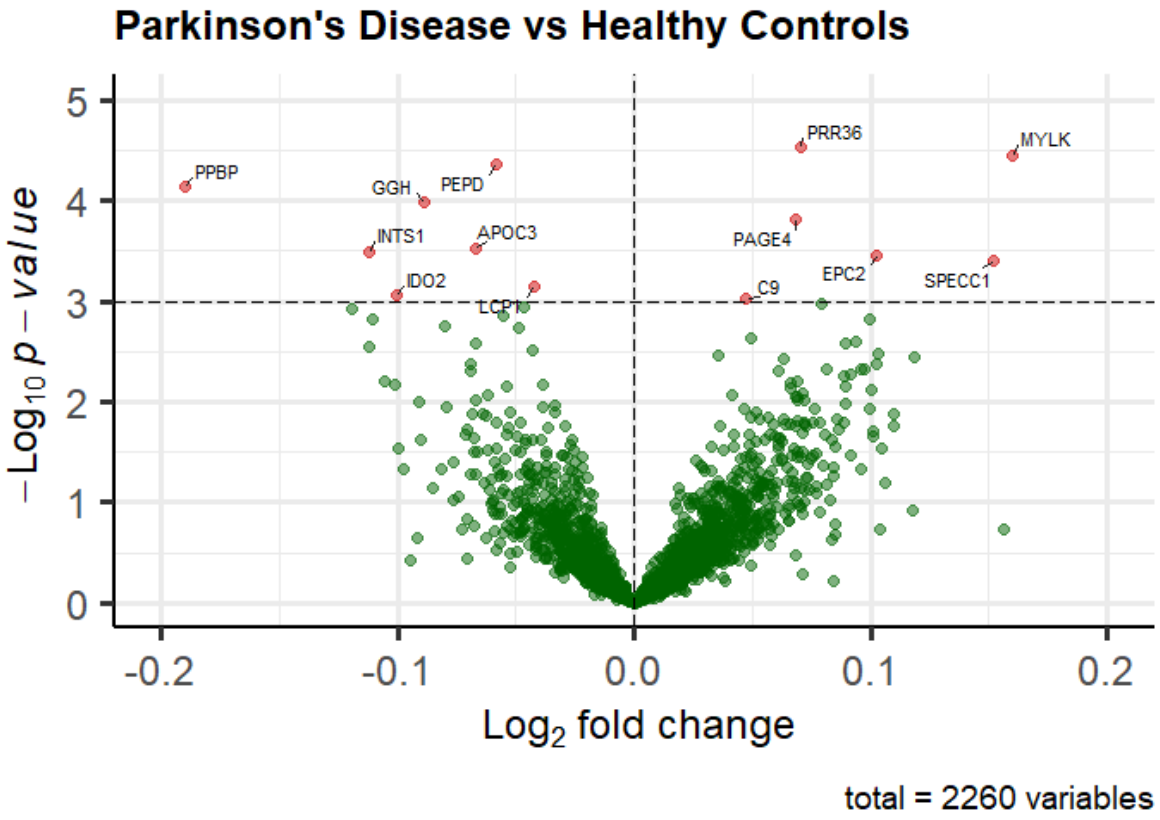

FIGURE 39 – VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE

Volcano plot of differentially expressed plasma proteins in Parkinson's disease versus Healthy Controls. Robust linear model generated -Log10 p-values on the y-axis, and log2 fold change on the x-axis. Proteins with p-value < 0.001 highlighted with their gene names.

TABLE 6 – TOP 30 DIFFERENTIALLY EXPRESSED PROTEINS FOR PARKINSON'S DISEASE VS HEALTHY CONTROLS

Top results from robust linear model of differentially expressed proteins in Parkinson's Disease versus Healthy Controls, ranked by p-value. Columns show Protein name, Gene name, Log2 Fold Change (logFC), average expression level of protein (AveExpr), t-statistic, p-value, Benjamini-Hochberg (BH) adjusted p-value, log odds B-statistic, and percentage of samples where the protein was quantified.

Protein name	*Gene*	*Log FC*	*Ave Expr*	*t-statistic*	*p-value*	*BH-adj. p-value*	*B-statistic*	*% Quantified*
Proline-rich protein 36	*PRR36*	0.070	2.621	4.275	0.000030	0.033	2.208	99
Myosin light chain kinase, smooth muscle	*MYLK*	0.160	2.193	4.239	0.000036	0.033	2.055	89
Xaa-Pro dipeptidase	*PEPD*	-0.058	2.687	-4.194	0.000044	0.033	1.875	86
Platelet basic protein	*PPBP*	-0.190	2.184	-4.083	0.000073	0.041	1.438	72
Gamma-glutamyl hydrolase	*GGH*	-0.089	2.062	-4.005	0.000104	0.047	1.132	64
P antigen family member 4	*PAGE4*	0.069	2.956	3.873	0.000154	0.058	0.738	82
Apolipoprotein C-III	*APOC3*	-0.067	3.143	-3.683	0.000296	0.086	0.098	100
Integrator complex subunit 1	*INTS1*	-0.112	1.945	-3.713	0.000321	0.086	0.147	55
Enhancer of polycomb homolog 2	*EPC2*	0.103	2.520	3.694	0.000343	0.086	0.089	55
Cytospin-B	*SPECC1*	0.152	1.719	3.639	0.000398	0.090	-0.052	61
Plastin-2	*LCP1*	-0.043	2.495	-3.441	0.000712	0.146	-0.682	95
Indoleamine 2,3-dioxygenase 2	*IDO2*	-0.101	2.280	-3.422	0.000870	0.161	-0.726	55
Complement component C9	*C9*	0.047	3.719	3.362	0.000924	0.161	-0.929	100
Terminal uridylyltransferase 7	*TUT7*	0.079	2.066	3.348	0.001049	0.166	-0.938	68
Proteoglycan 4	*PRG4*	-0.047	3.518	-3.302	0.001135	0.166	-1.118	100

1-phosphatidylinositol 3-phosphate 5-kinase	*PIKFYVE*	-0.119	2.818	-3.295	0.001178	0.166	-1.086	92
Basic helix-loop-helix domain-containing protein USF3	*USF3*	-0.055	2.130	-3.260	0.001402	0.180	-1.183	68
Fez family zinc finger protein 2	*FEZF2*	0.100	1.762	3.249	0.001483	0.180	-1.213	62
TLC domain-containing protein 4	*TLCD4*	-0.110	2.493	-3.253	0.001512	0.180	-1.213	55
Astrotactin-2	*ASTN2*	-0.080	2.196	-3.200	0.001754	0.194	-1.349	59
Prenylcysteine oxidase 1	*PCYOX1*	-0.049	2.853	-3.164	0.001798	0.194	-1.526	100
Epiplakin	*EPPK1*	0.050	2.442	3.095	0.002313	0.238	-1.681	82
Nucleoplasmin-2	*NPM2*	0.094	2.148	3.056	0.002561	0.241	-1.828	95
Eukaryotic translation initiation factor 3 subunit L	*EIF3L*	-0.067	2.208	-3.068	0.002656	0.241	-1.713	59
Immunoglobulin lambda variable 5-39	*IGLV5-39*	0.090	2.364	3.046	0.002663	0.241	-1.845	90
Thrombospondin-1	*THBS1*	-0.112	1.981	-3.035	0.002816	0.245	-1.825	77
L-selectin	*SELL*	-0.043	2.697	-2.994	0.003117	0.261	-1.998	95
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	*MGAT4A*	0.103	2.430	2.994	0.003349	0.270	-1.908	59
Putative alpha-1-antitrypsin-related protein	*SERPINA2*	0.036	3.742	2.954	0.003509	0.273	-2.113	99
Integrin alpha-6	*ITGA6*	0.118	2.492	2.967	0.003627	0.273	-1.982	59

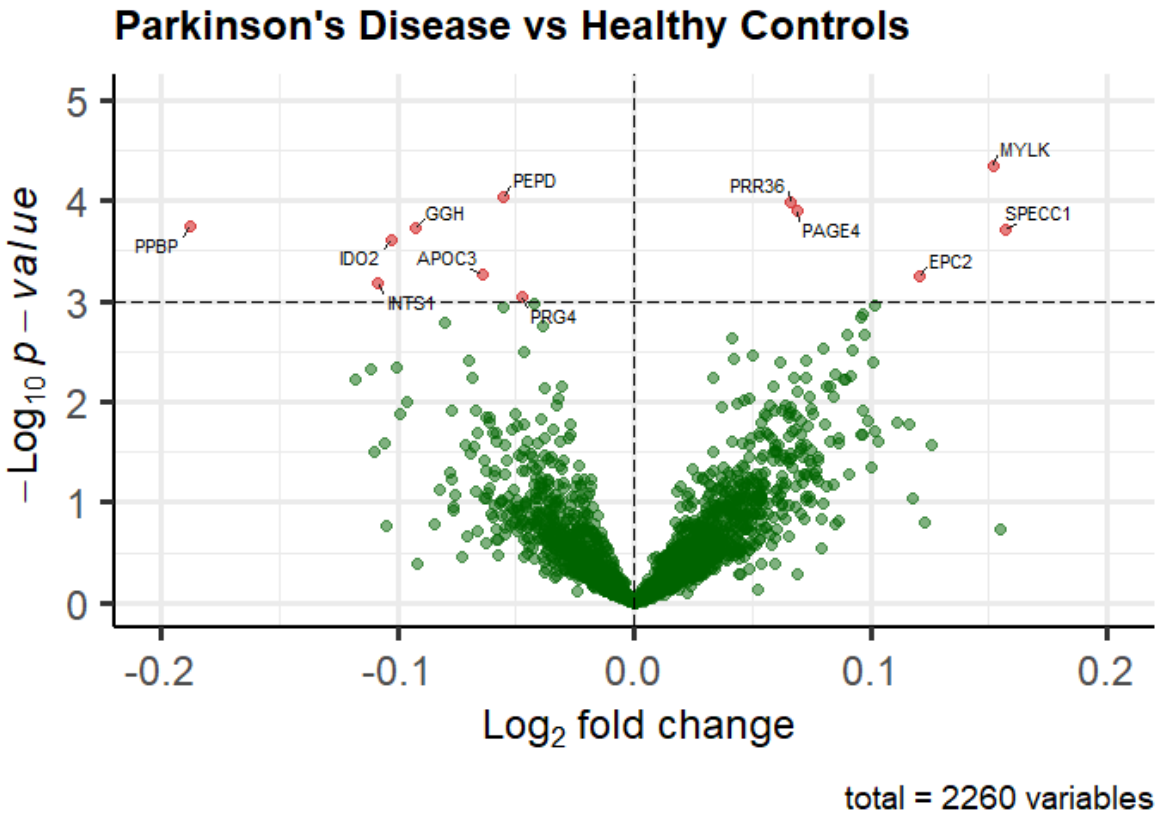

FIGURE 40 – VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE – AGE AND SEX ADJUSTED

Volcano plot of differentially expressed plasma proteins in Parkinson's disease versus Healthy Controls, with age and sex as added covariates. Robust linear model generated -Log10 p-values on the y-axis, and log2 fold change on the x-axis. Proteins with p-value < 0.001 highlighted with their gene names.

TABLE 7 – TOP 30 DIFFERENTIALLY EXPRESSED PROTEINS FOR PARKINSON'S DISEASE VS HEALTHY CONTROLS

Top results from robust linear model of differentially expressed proteins in Parkinson's Disease versus Healthy Controls, with age and sex as covariates, ranked by p-value. Columns show Protein name, Gene name, Log2 Fold Change (logFC), average expression level of protein (AveExpr), t-statistic, p-value, Benjamini-Hochberg (BH) adjusted p-value, log odds B-statistic, and percentage of samples where the protein was quantified.

Protein name	*Gene*	*logFC*	*AveExpr*	*t-statistic*	*p-value*	*BH-adj. p-value*	*B-statistic*	*% quantified*
Myosin light chain kinase, smooth muscle	*MYLK*	0.152	2.193	4.185	0.000045	0.063	1.846	89
Xaa-Pro dipeptidase	*PEPD*	-0.055	2.687	-4.012	0.000090	0.063	1.237	86
Proline-rich protein 36	*PRR36*	0.066	2.621	3.967	0.000101	0.063	1.073	99
P antigen family member 4	*PAGE4*	0.069	2.956	3.931	0.000124	0.063	0.940	82
Platelet basic protein	*PPBP*	-0.188	2.184	-3.844	0.000180	0.063	0.627	72
Gamma-glutamyl hydrolase	*GGH*	-0.092	2.062	-3.850	0.000186	0.063	0.620	64
Cytospin-B	*SPECC1*	0.157	1.719	3.841	0.000195	0.063	0.575	61
Indoleamine 2,3-dioxygenase 2	*IDO2*	-0.103	2.280	-3.792	0.000245	0.069	0.379	55
Apolipoprotein C-III	*APOC3*	-0.064	3.143	-3.515	0.000544	0.128	-0.454	100
Enhancer of polycomb homolog 2	*EPC2*	0.121	2.520	3.552	0.000565	0.128	-0.342	55
Integrator complex subunit 1	*INTS1*	-0.108	1.945	-3.508	0.000656	0.135	-0.492	55
Proteoglycan 4	*PRG4*	-0.048	3.518	-3.372	0.000895	0.169	-0.906	100
Plastin-2	*LCP1*	-0.042	2.495	-3.329	0.001046	0.174	-1.029	95
Fez family zinc finger protein 2	*FEZF2*	0.102	1.762	3.344	0.001090	0.174	-0.962	62
Basic helix-loop-helix domain-containing protein USF3	*USF3*	-0.055	2.130	-3.319	0.001157	0.174	-1.028	68
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	*MGAT4A*	0.097	2.430	3.291	0.001317	0.186	-1.144	59

Transmembrane protease serine 6	*TMPRSS6*	0.096	2.960	3.251	0.001472	0.196	-1.255	64
Astrotactin-2	*ASTN2*	-0.080	2.196	-3.223	0.001640	0.206	-1.313	59
L-selectin	*SELL*	-0.039	2.697	-3.178	0.001730	0.206	-1.462	95
Immunoglobulin lambda variable 5-39	*IGLV5-39*	0.090	2.364	3.116	0.002131	0.234	-1.656	90
Rho guanine nucleotide exchange factor 17	*ARHGEF17*	0.097	2.690	3.112	0.002176	0.234	-1.612	86
Complement component C9	*C9*	0.042	3.719	3.084	0.002333	0.240	-1.750	100
Terminal uridylyltransferase 7	*TUT7*	0.080	2.066	3.027	0.002952	0.283	-1.821	68
AP-1 complex subunit gamma-1	*AP1G1*	0.092	2.852	3.012	0.003005	0.283	-1.921	82
Prenylcysteine oxidase 1	*PCYOX1*	-0.047	2.853	-2.984	0.003199	0.289	-2.041	100
Epiplakin	*EPPK1*	0.050	2.442	2.968	0.003448	0.298	-2.004	82
Uveal autoantigen with coiled-coil domains and ankyrin repeats	*UACA*	0.042	2.459	2.945	0.003657	0.298	-2.115	90
Zinc finger homeobox protein 4	*ZFHX4*	0.072	2.425	2.931	0.003856	0.298	-2.133	82
G patch domain-containing protein 8	*GPATCH8*	-0.070	2.204	-2.947	0.003930	0.298	-2.051	54
Complement decay-accelerating factor	*CD55*	0.101	1.681	2.921	0.004049	0.298	-2.123	72

The 10 most significantly differentially expressed proteins from the age and sex corrected analysis are shown in Figure 41. These include Myosin Light Chain Kinase, smooth muscle (MYLK), Xaa-Pro dipeptidase (PEPD), Proline-rich protein 36 (PRR36), P antigen family member 4 (PAGE4), Platelet basic protein (PPBP), Gamma-glutamyl hydrolase (GGH), Cytospin-B (SPECC1), Indoleamine 2,3-dioxygenase 2 (IDO2), Apolipoprotein C-III (APOC3), and Enhancer of polycomb homolog 2 (EPC2). It was seen that there indeed was a highly significant group difference, however there was still a large group overlap present.

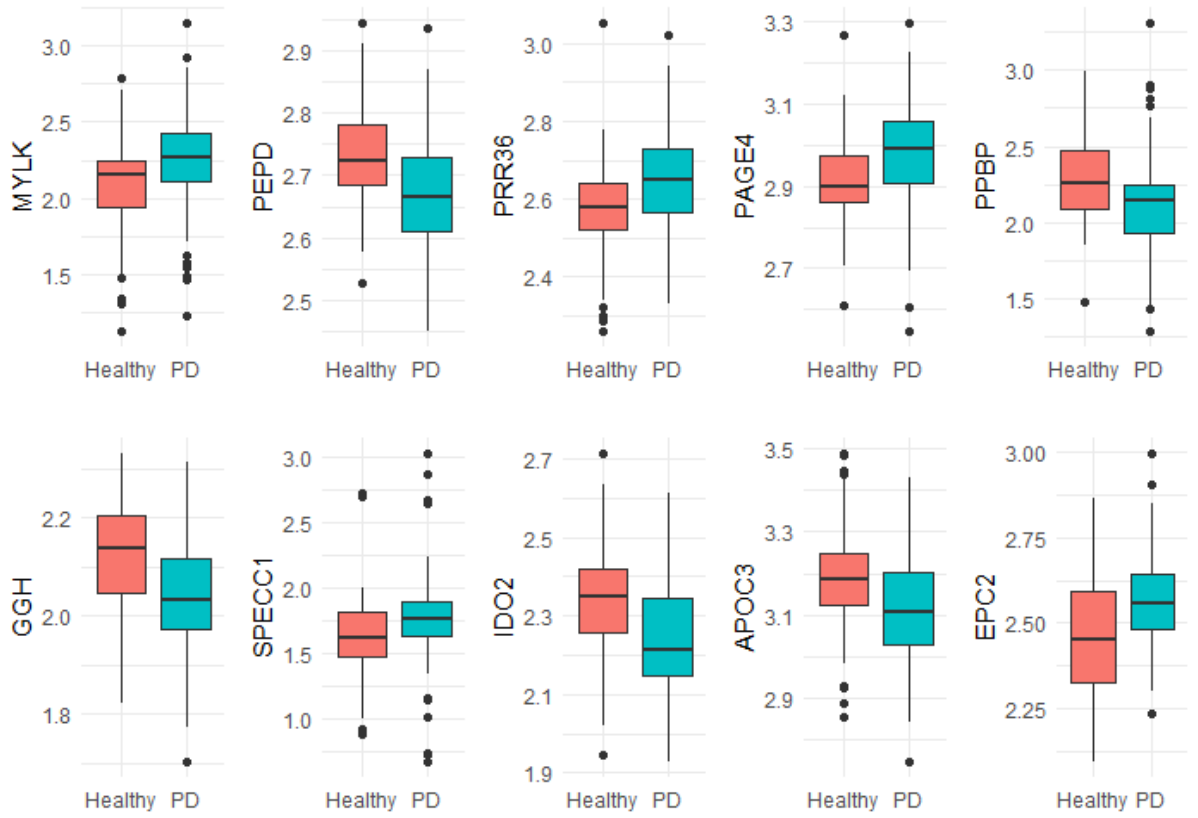

FIGURE 41 – BOXPLOTS OF TOP 10 HITS – PARKINSON'S DISEASE VERSUS HEALTHY CONTROLS

Boxplots of the top 10 most significant differentially expressed proteins between Parkinson's Disease (PD) and healthy controls, adjusted for age and sex. Y-axis represents log transformed relative protein abundance. Gene names presented; Myosin Light Chain Kinase, smooth muscle (MYLK), Xaa-Pro dipeptidase (PEPD), Proline-rich protein 36 (PRR36), P antigen family member 4 (PAGE4), Platelet basic protein (PPBP), Gamma-glutamyl hydrolase (GGH), Cytospin-B (SPECC1), Indoleamine 2,3-dioxygenase 2 (IDO2), Apolipoprotein C-III (APOC3), and Enhancer of polycomb homolog 2 (EPC2).

Although ranking the proteins by p-value as the primary outcome statistic is useful to determine the confidence of the protein truly being differentially expressed, it does not necessarily reflect which protein is most differentially expressed. Ranking the proteins by effect size, which is independent from sample size, may prove to be a more useful strategy to identify the best candidate markers. Given that the sample distribution around the mean, and therefore standard deviation, varies between proteins, log fold change is not an ideal measurement of effect size. A protein with a large standard deviation would show a greater group overlap than a protein with a smaller standard deviation but with the same fold change. Therefore, Cohen's d effect size was calculated, which is dependent on distribution of each sample as well as the fold change, but not on the sample size. The top hits are displayed in

figure 42 and Table 8. It was noted that some of the most significant proteins where generally lower percentage of samples were quantified had a higher effect size. This makes sense as the confidence in rejecting the null hypothesis decreases with sample size, irrespective of the effect size. When the absolute value of Cohen's d was set to >0.6, 17 candidates were detected. These hits were largely the same as when filtering by p-value but ranked in a different order.

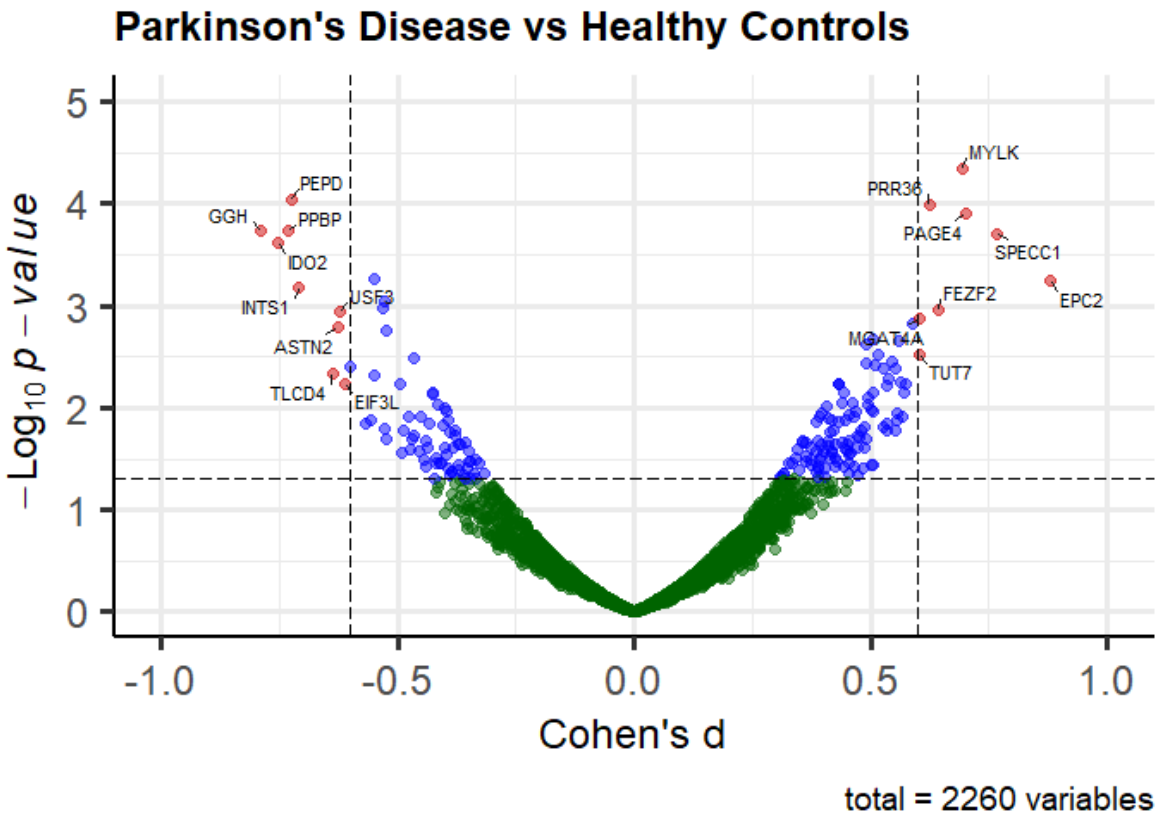

FIGURE 42 – VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE – AGE AND SEX ADJUSTED WITH EFFECT SIZE CUTOFF

Volcano plot of differentially expressed plasma proteins in Parkinson's Disease versus Healthy Controls, with age and sex as added covariates. Robust linear model generated -Log10 p-values on the y-axis, and Cohen's d effect size on the x-axis. Proteins with p-value < 0.05, and |Cohen's d|>0.6 highlighted with their gene names.

TABLE 8 –TOP DIFFERENTIALLY EXPRESSED PROTEINS FOR PARKINSON'S DISEASE VS HEALTHY CONTROLS – BY EFFECT SIZE

Top results from robust linear model of differentially expressed proteins in Parkinson's Disease versus Healthy Controls, with age and sex as covariates. Columns show Protein name, Gene name, Log2 Fold Change (logFC), average expression level of protein (AveExpr), t-statistic, p-value, and percentage of samples where the protein was quantified, and Cohen's d effect size. Significantly differentiated proteins with |Cohen's d|>0.6 displayed.

Protein name	Gene	logFC	AveExpr	t-statistic	p-value	% quantified	Cohen's d
Enhancer of polycomb homolog 2	EPC2	0.121	2.520	3.552	0.00057	55	0.881
Gamma-glutamyl hydrolase	GGH	-0.092	2.062	-3.850	0.00019	64	-0.789
Cytospin-B	SPECC1	0.157	1.719	3.841	0.00019	61	0.767
Indoleamine 2,3-dioxygenase 2	IDO2	-0.103	2.280	-3.792	0.00024	55	-0.754
Platelet basic protein	PPBP	-0.188	2.184	-3.844	0.00018	72	-0.733
Xaa-Pro dipeptidase	PEPD	-0.055	2.687	-4.012	0.00009	86	-0.726
Integrator complex subunit 1	INTS1	-0.108	1.945	-3.508	0.00066	55	-0.709
P antigen family member 4	PAGE4	0.069	2.956	3.931	0.00012	82	0.701
Myosin light chain kinase, smooth muscle	MYLK	0.152	2.193	4.185	0.00004	89	0.692
Fez family zinc finger protein 2	FEZF2	0.102	1.762	3.344	0.00109	62	0.642
TLC domain-containing protein 4	TLCD4	-0.101	2.493	-2.894	0.00459	55	-0.639
Astrotactin-2	ASTN2	-0.080	2.196	-3.223	0.00164	59	-0.628
Proline-rich protein 36	PRR36	0.066	2.621	3.967	0.00010	99	0.623
Basic helix-loop-helix domain-containing protein USF3	USF3	-0.055	2.130	-3.319	0.00116	68	-0.622
Eukaryotic translation initiation factor 3 subunit L	EIF3L	-0.068	2.208	-2.812	0.00575	59	-0.611
Alpha-1,3-mannosyl-glycoprotein 4-beta-N-acetylglucosaminyltransferase A	MGAT4A	0.097	2.430	3.291	0.00132	59	0.605
Terminal uridylyltransferase 7	TUT7	0.080	2.066	3.027	0.00295	68	0.604

3.1.4.2.3 PEPTIDE ANALYSIS

As an additional exploratory step, individual peptides were analysed separately in order to investigate whether certain peptides were more differentially expressed and would perform better as biomarkers than others. The peptides were pre-processed in the same way as the proteins by normalising the data, correcting for intra-run quality controls, and corrected for other technical variation. Each protein is composed of many different peptides, and often different sets of peptides are detected in the different samples that are summed up to protein abundances. Therefore, setting the threshold for each peptide to be quantified in at least 50% of samples, as was done with the protein analysis, eliminated most peptides that did not belong to highly abundant plasma proteins. Therefore, it was only required that the peptides were detected in at least 25% of the samples. The peptides were analysed using a linear model with age and sex as covariates (Figure 43).

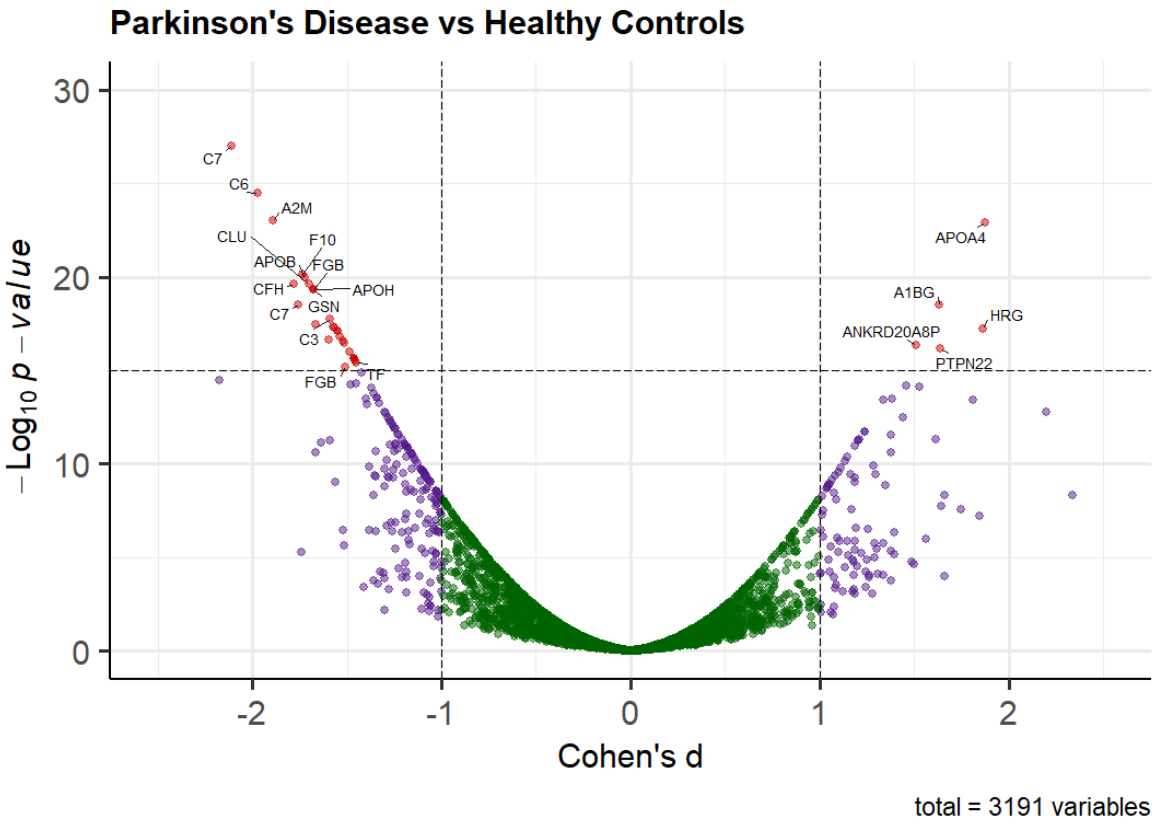

FIGURE 43 - VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PEPTIDES IN PARKINSON'S DISEASE – AGE AND SEX ADJUSTED WITH P-VALUE CUTOFF

Volcano plot of differentially expressed plasma peptides in Parkinson's Disease versus Healthy Controls, with age and sex as added covariates. Linear model generated -Log10 p-values on the y-axis, and Cohen's d effect size on the x-axis. Proteins with -log10 BH adj. p-value > 30, and |Cohen's d| > 1.0 highlighted with their gene names.

Several peptides were significantly differentially expressed in PD versus HC. Most of the peptides belonged to the most abundant protein groups in plasma, such as apolipoproteins, coagulation factors, or complement proteins. Hence, the p-value cut-off was set to 0.05 instead, with a more stringent cut-off for effect size (|Cohen's d| > 1.5) to identify the lower abundant differentially expressed plasma proteins in PD (Figure 44).

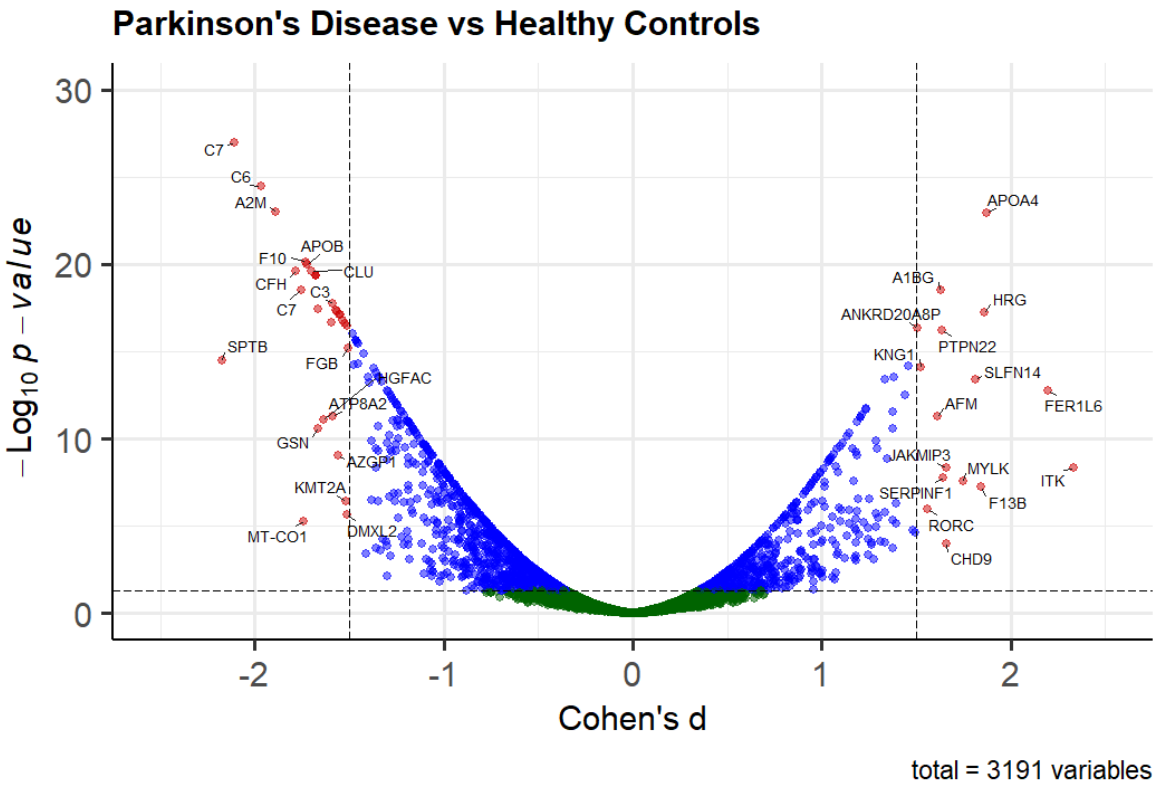

FIGURE 44- VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PEPTIDES IN PARKINSON'S DISEASE – AGE AND SEX ADJUSTED WITH EFFECT SIZE CUTOFF

Volcano plot of differentially expressed plasma peptides in Parkinson's Disease versus Healthy Controls, with age and sex as added covariates. Linear model generated -Log10 p-values on the y-axis, and Cohen's d effect size on the x-axis. Proteins with BH adj. p-value < 0.05, and |Cohen's d| > 1.5 highlighted with their gene names.

When sorting by effect size, peptides from several lower abundant proteins were found. For example, the KLPAENGSSSAETLNAK peptide of MYLK, which one of the most significantly changed proteins in PD, seemed to be one of the most consistently elevated peptides in PD. However, the peptide was still only quantified in 35% of the samples. A total of 233 protein unique peptides were

used for the protein quantification of MYLK, most peptides were only quantified in a few samples, and only 2 peptides were individually quantified in >25% of the samples.

Finally, only peptides from the proteins that were significantly changed in PD ($p < 0.05$), and quantified in at least 10% of the samples were analysed. 427 peptides from 138 protein groups were analysed (Figure 45), and the top results are displayed in Table 9, sorted by effect size. The peptide with the largest effect size belonged to Eukaryotic Translation Initiation Factor 3 Subunit L (EIF3L) with Cohen's d = 2.50. Again, most peptides were detected in a minority of the samples, with the exception of high abundant proteins such as complement factors C7 and C9.

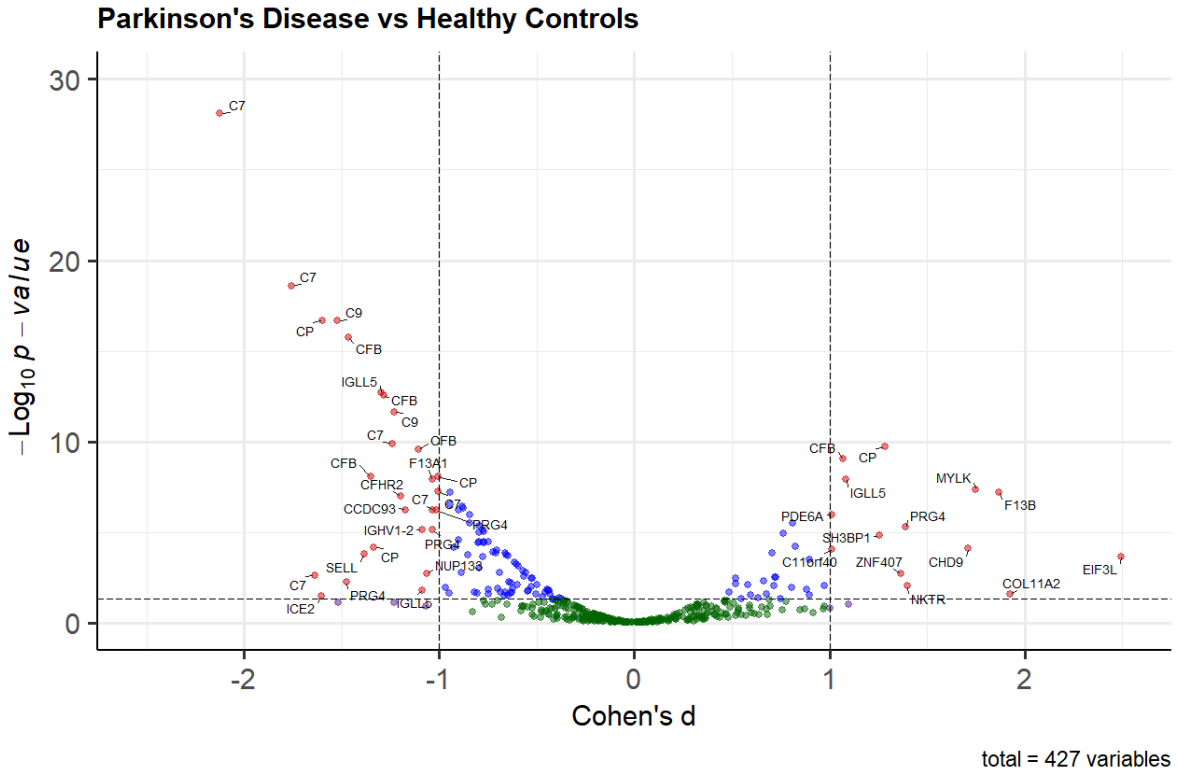

FIGURE 45 – VOLCANO PLOT OF PEPTIDES FROM DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE VS HEALTHY CONTROLS

Volcano plot of differentially expressed plasma peptides in Parkinson's Disease versus Healthy Controls, with age and sex as added covariates. Only peptides from significantly changed proteins in PD plasma were analysed. Linear model generated -Log10 p-values on the y-axis, and Cohen's d effect size on the x-axis. Proteins with BH adj. p-value < 0.05, and |Cohen's d| > 1.0 highlighted with their gene names.

TABLE 9 –PEPTIDES FROM DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE VS HEALTHY CONTROLS – BY EFFECT SIZE

Analysis of individual peptides from significantly changed proteins in PD plasma. Top results from robust linear model of differentially expressed peptides in Parkinson's Disease versus Healthy Controls, with age and sex as covariates. Columns show Gene name, Peptide sequence, Benjamini-Hochberg corrected p-value, percentage of samples where the peptides were quantified, and Cohen's d effect size. Top 30 Significantly differentiated peptides ordered by effect size.

Gene	*Peptide sequence*	*Adj. p-value*	*% Quantified*	*Cohen's d*
EIF3L	VFSDEVQQQAQLSTIRSFLKLYTTMPVAK	0.000220847	11.61616162	2.496198199
C7	CDAESSK	7.77217E-29	99.49494949	-2.123050339
COL11A2	GAKGATGPGGPKGEK	0.026137504	10.1010101	1.924917658
F13B	LSFFCLAGYTTESGRQEEQTTCTTEGWSPEPR	6.07447E-08	29.29292929	1.868489907
C7	ECNNPPPSGGGR	2.45904E-19	90.4040404	-1.75544827
MYLK	KLPAENGSSSAETLNAK	4.35767E-08	35.35353535	1.747391055
CHD9	KVGGAFAPPLK	8.27709E-05	26.76767677	1.709918771
C7	GQSISVTSIRPCAAETQ	0.002406416	15.15151515	-1.638640948
ICE2	ASDGKVTRTAYNLYK	0.03428823	12.12121212	-1.602609935
CP	TYSDHPEKVNKDDEEFIESNK	2.08132E-17	90.90909091	-1.596863647
C9	VVEESELAR	2.08132E-17	99.49494949	-1.520640714
PRG4	CGEGYSR	0.005733871	15.65656566	-1.475531827
CFB	LEDSVTYHCSR	1.67992E-16	100	-1.46540719
NKTR	LDTPDINIVLK	0.009580989	15.65656566	1.397097296
PRG4	EPAPTTPKEPAPTTTKKPAPTTPK	5.19434E-06	32.82828283	1.391347482
SELL	WNDDACHK	0.000169712	29.29292929	-1.386146693
ZNF407	HLGMREYK	0.001860845	19.6969697	1.365116735
CFB	LPPTTTCQQQK	8.67319E-09	68.18181818	-1.352850094
CP	MYYSAVDPTK	6.70056E-05	38.38383838	-1.335203227
IGLL5	TVAPTECS	2.06955E-13	99.49494949	-1.298089865
CFB	VSEADSSNADWVTK	2.67817E-13	100	-1.285875575
CP	GPEEEHLGILGPVIWAEVGDTIR	1.94471E-10	81.81818182	1.282625918
SH3BP1	SRLSQATK	1.56897E-05	34.34343434	1.253965958
C7	MPYECGPSLDVCAQDER	1.30158E-10	82.82828283	-1.242324409
C9	CTDAVGDR	2.28012E-12	99.49494949	-1.23310262
CFHR2	NGQWSEPPK	9.789E-08	68.18181818	-1.19817159
CCDC93	TVVDLSEVYKPR	5.87958E-07	59.5959596	-1.172050595
CFB	DAQYAPGYDK	2.57138E-10	99.49494949	-1.106378919
IGHV1-2	SDDTAVYYCAR	7.45427E-06	63.63636364	-1.08966516
IGLL5	VTVLGQPK	0.016626075	17.67676768	-1.085645607

3.1.4.2.4 SEX-SPECIFIC ANALYSIS

The dataset was sex matched between PD and controls, however there were more males (n = 107) than females (n = 91) in the cohort. Indeed the incidence of PD in males is higher than in females, and sex-specific phenotypes and biomarkers have previously been described in PD [15, 193]. Hence, the proteomic data was additionally separately analysed for males and females to investigate whether there were any sex related differences. Similar to the analysis of the full data set, the male and female only datasets were adjusted for technical group effects, and proteins that were quantified in at least 50% of the samples were selected for the analysis. The differential protein expression was more pronounced in the sex-specific analysis, particularly in regards to the effect size. A higher effect size cut-off (Cohen's d > 0.75) was set for volcano plots for the sex-specific analyses. The results are presented in the volcano plots for PD versus HC in females (Figure 46) and males (Figure 47) respectively, and the statistics are summarised in Table 10.

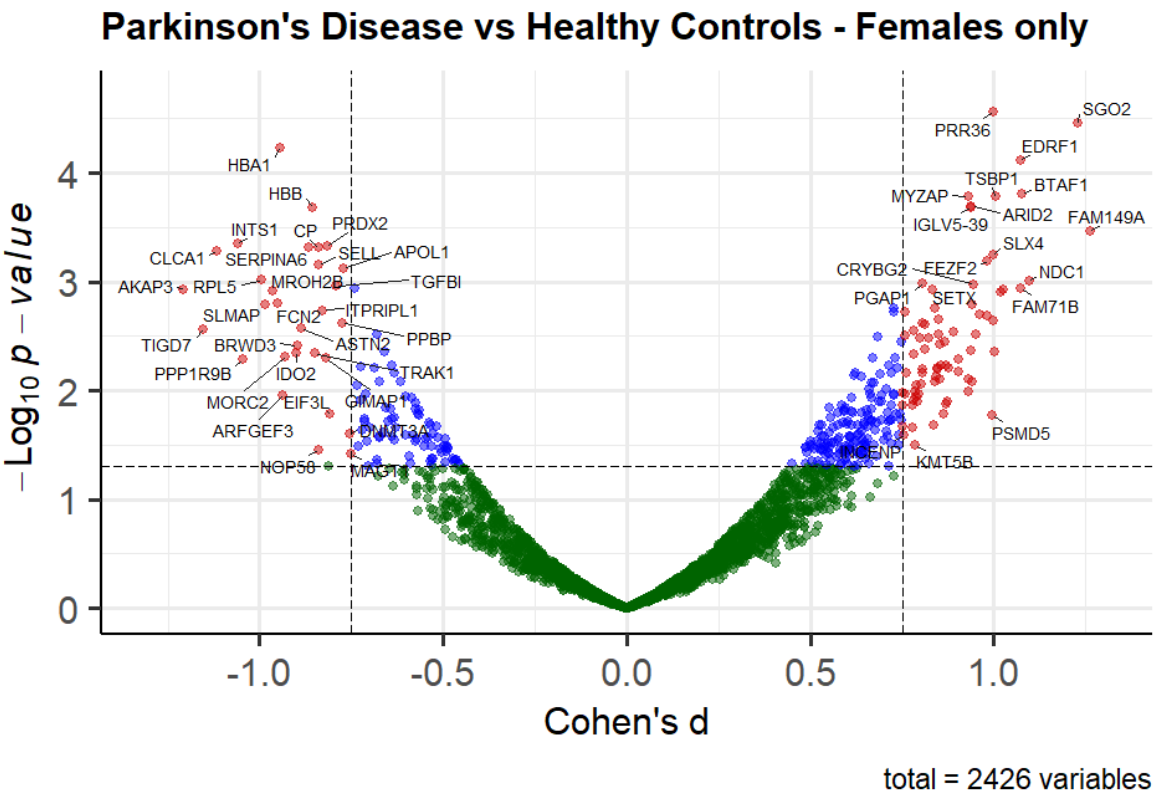

FIGURE 46 - VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE – FEMALES ONLY WITH EFFECT SIZE CUT-OFF

Volcano plot of differentially expressed plasma proteins in female Parkinson's Disease patients versus female Healthy Controls, with age as a covariate. Robust linear model generated -Log10 p-values on the y-axis, and Cohen's d effect size on the x-axis. Proteins with p-value < 0.05, and |Cohen's d| > 0.75 highlighted with their gene names.

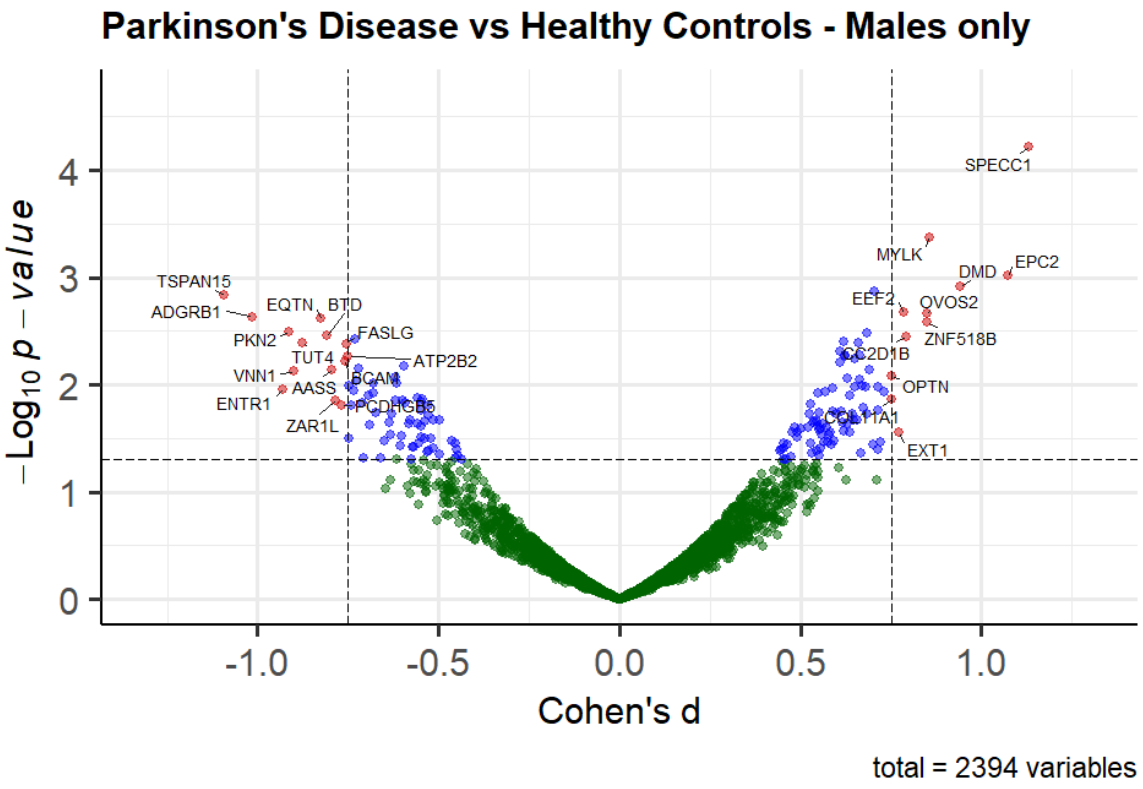

FIGURE 47 - VOLCANO PLOT OF DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE – MALES ONLY WITH EFFECT SIZE CUT-OFF

Volcano plot of differentially expressed plasma proteins in male Parkinson's Disease patients versus male Healthy Controls, with age as a covariate. Robust linear model generated -Log10 p-values on the y-axis, and Cohen's d effect size on the x-axis. Proteins with p-value < 0.05, and |Cohen's d| > 0.75 highlighted with their gene names.

TABLE 10 –TOP DIFFERENTIALLY EXPRESSED PROTEINS FOR PARKINSON'S DISEASE VS HEALTHY CONTROLS – BY EFFECT SIZE – SEPARATED IN MALES AND FEMALES

Top results from robust linear model of differentially expressed proteins in Parkinson's Disease versus Healthy Controls, with age as covariate, males and females analysed separately. Columns Gene name, p-value, number of samples where the protein was quantified, and Cohen's d effect size. Top 30 proteins ordered by effect size for females and males respectively.

Females				Males			
Gene	*p-value*	*n quantified*	*Cohen's d*	*Gene*	*p-value*	*n quantified*	*Cohen's d*
FAM149A	0.00035	48	1.262	*SPECC1*	0.00006	65	1.131
SGO2	0.00003	67	1.228	*TSPAN15*	0.00146	58	-1.092
AKAP3	0.00119	47	-1.206	*EPC2*	0.00096	57	1.074
TIGD7	0.00273	46	-1.152	*ADGRB1*	0.00233	56	-1.013
CLCA1	0.00053	55	-1.117	*DMD*	0.00120	68	0.942
NDC1	0.00100	49	1.096	*ENTR1*	0.01112	57	-0.930
BTAF1	0.00016	69	1.077	*PKN2*	0.00319	63	-0.915
FAM71B	0.00116	51	1.073	*VNN1*	0.00747	58	-0.900
EDRF1	0.00008	70	1.072	*TUT4*	0.00411	57	-0.875
INTS1	0.00045	47	-1.060	*MYLK*	0.00042	97	0.855
PPP1R9B	0.00521	46	-1.046	*ZNF518B*	0.00258	54	0.850
LRRC9	0.00118	52	1.025	*OVOS2*	0.00214	63	0.848
F2RL1	0.00125	51	1.020	*EQTN*	0.00243	74	-0.825
TSBP1	0.00017	64	1.004	*BTD*	0.00346	69	-0.810
ITGA6	0.00444	52	1.000	*AASS*	0.00723	58	-0.795
EPB41L2	0.00228	46	0.998	*CC2D1B*	0.00353	78	0.792
PRR36	0.00003	91	0.998	*EEF2*	0.00210	85	0.785
SLX4	0.00057	60	0.997	*ZAR1L*	0.01398	54	-0.783
PSMD5	0.01704	48	0.995	*EXT1*	0.02810	54	0.773
RPL5	0.00096	51	-0.993	*PCDHGB5*	0.01542	70	-0.769
SLMAP	0.00162	46	-0.983	*BCAM*	0.00599	71	-0.759
FEZF2	0.00064	59	0.983	*FASLG*	0.00412	67	-0.755
PLXNB1	0.00205	47	0.981	*COL11A1*	0.01377	55	0.753
MROH2B	0.00121	57	-0.964	*OPTN*	0.00819	66	0.752
BAHCC1	0.00200	55	0.962	*ATP2B2*	0.00549	68	-0.750
FCN2	0.00159	62	-0.951	*COL1A1*	0.01033	57	-0.748
PEX3	0.00303	54	0.950	*SPECC1L*	0.03173	60	-0.747
CRYBG2	0.00106	68	0.945	*ZNF750*	0.01546	67	-0.740
HBA1	0.00006	91	-0.943	*PIP5K1C*	0.01137	59	-0.735
ZC3H14	0.00836	57	0.942	*IGKV2-29*	0.01153	69	0.730

Interestingly, female patients had more differentially expressed proteins than male patients. P-value histograms from the analyses confirmed significantly more changed proteins in the female cohort compared with the male cohort (Figure 48). Interestingly, most of the significantly changed proteins in the two analyses did not overlap, which suggest many of the differentially expressed proteins are sex-specific. It was theorised that the lower number of altered proteins in the male cohort was due to a more heterogeneous PD population. Therefore, a few demographic variables were compared using density plots between the male and female PD cohort (Figure 49). No apparent major demographic differences were observed between the sexes. Statistical analysis with Wilcoxon rank test revealed significantly higher Hoehn & Yahr disease stage for females (p=0.042), and higher LEDD scores for males (p=0.038).

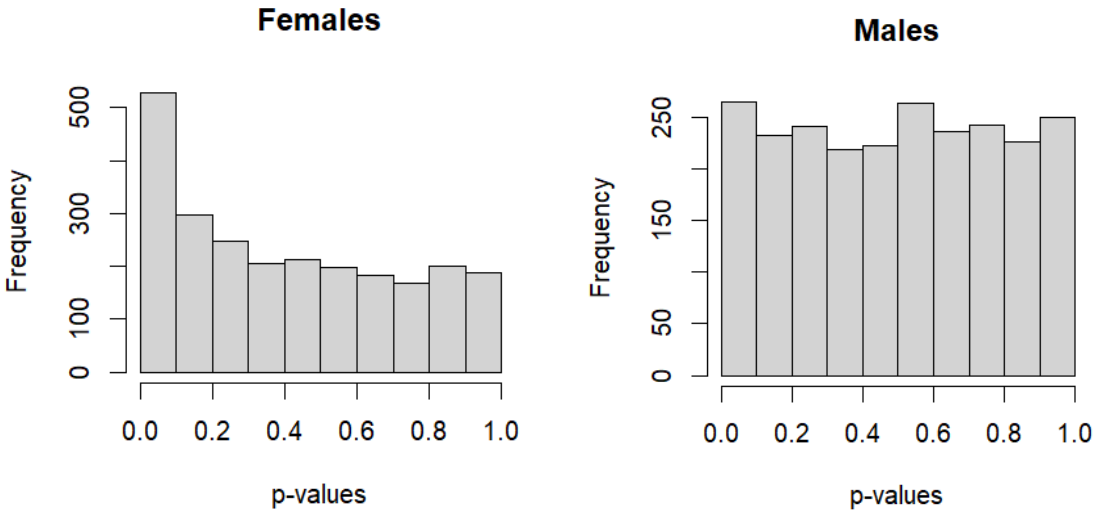

FIGURE 48 – P-VALUE HISTOGRAM FROM THE GROUPWISE COMPARISON OF PROTEINS BETWEEN PARKINSON'S DISEASE AND HEALTHY CONTROLS – SPLIT INTO MALES AND FEMALES

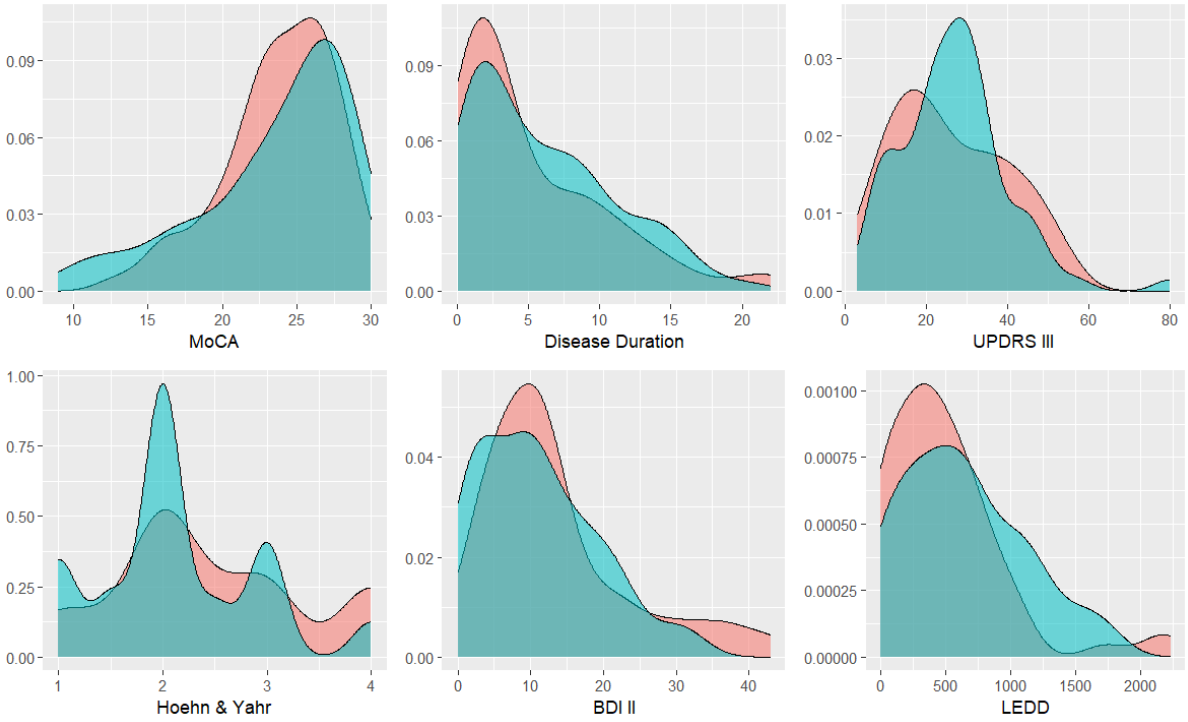

FIGURE 49 – DENSITY PLOTS OF DEMOGRAPHIC AND CLINICAL VARIABLES IN MALE (BLUE) AND FEMALE (RED) PATIENTS WITH PARKINSON'S DISEASE

Distributions of Montreal Cognitive Assessment (MoCA), Disease Duration in years, Unified Parkinson's Disease Rating Scale (UPDRS) part III, Hoehn & Yahr, Beck Depression Inventory II (BDI II), and Levodopa Equivalent Daily Dose (LEDD) are displayed.

Following these results, the effect of cohort heterogeneity was further explored. When more severe PD cases were excluded from the analysis, generally the proportion of significantly altered proteins between PD and HC increased. This was assessed though the proportion of proteins with a p-value < 0.05 in the p-value histogram. Excluding patients with more severe motor symptoms, high levodopa treatment, long disease duration or low cognitive score all yielded similar results with an increased proportion of significantly changed proteins. It was not surprising eliminating severely affected PD patients using any of these variables had a similar effect on the results, given that these variables are collinearly explaining the variation in the cohort as seen in the PCA plot in Figure 20. Furthermore, this resulted in an increase in significantly differential proteins for both the male and female cohort. As an example, the p-value histograms in Figure 50 shows a larger number of true positive hits after selecting PD patients without any cognitive impairment. The trade-off from these analyses is the much smaller cohort size, and the increased risk of false positive results.

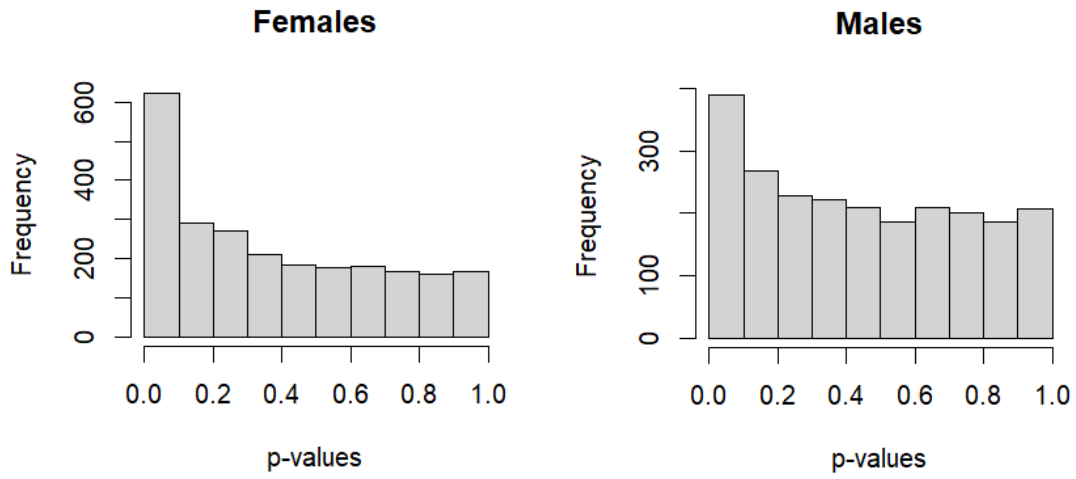

FIGURE 50 – P-VALUE HISTOGRAM FROM THE GROUPWISE COMPARISON OF PROTEINS BETWEEN COGNITIVELY UNIMPAIRED PARKINSON'S DISEASE AND HEALTHY CONTROLS – SPLIT INTO FEMALES AND MALES

3.1.4.2.5 BIOMARKER PANELS AND MACHINE LEARNING

Calculating the confidence of a differentially expressed protein with a p-value, or how large the effect size is giving a good idea of whether the protein is a candidate biomarker.

However, the best diagnostic biomarker for a disease is not necessarily determined with single parametric group comparisons. Ideally from a performance perspective, a biomarker should be resistant to demographic variables and show consistent discriminatory performance when the model is applied to new data. In brief, a good biomarker can perform predictions consistently. Therefore, a machine learning algorithm was trained for diagnosis prediction, and applied on a subset of data.

Proteins with unadjusted p-values <0.05 from the robust linear model above were used in the analysis. 90% of the data was selected at random as training data a total of 5 times using bootstrapping. A glm logistic regression model was constructed with diagnosis as the outcome variable, and protein quantification, age, and sex as explanatory variables. The glm model was in turn trained on itself using a leave one out cross validation method. The final model was subsequently applied on the test data. The average accuracy and AUC were recorded, as well as the importance of age, sex, and protein quantification, respectively. A separate analysis was made on a covariate free model, where accuracy and AUC was also recorded. A threshold was set on minimum 65% AUC and a relative importance of the protein of >50%. Results are presented in Table 11.

TABLE 11 – TOP DIFFERENTIALLY EXPRESSED PROTEINS FOR PARKINSON'S DISEASE VS HEALTHY CONTROLS – BY CROSS VALIDATED PREDICTIVE PERFORMANCE

Top results from robust linear model of differentially expressed proteins in Parkinson's Disease versus Healthy Controls, with age and sex as covariates with. Cohen's d, p-value, and % of samples where the protein was quantified are listed. In addition, proteins with unadjusted p-values <0.05 were used to build glm logistic regression model with and protein levels, age, and sex as explanatory variables. The glm model was trained on itself and tested on a subset of data. The average accuracy and area under curve (AUC) are presented, as well as the importance of age, sex, and protein as variables. A separate analysis was made on a covariate free model, where accuracy and AUC was also recorded, represented in the rightmost two columns. Proteins displayed had a minimum 65% AUC and a relative importance of the protein of >50%.

Gene	*Cohen's d*	*p-value*	*% quant*	*imp_ Age*	*imp_ Sex*	*imp_ Prot*	*ACC all var*	*AUC all var*	*ACC prot*	*AUC prot*
F2RL1	0.56	0.00556	58	0.39	0.23	2.43	0.76	0.69	0.71	0.71
INTS1	-0.71	0.00066	55	1.07	0.24	3.03	0.74	0.72	0.65	0.64
CP	-0.43	0.00699	100	1.82	0.28	2.35	0.74	0.76	0.66	0.62
PRR36	0.62	0.00010	99	0.76	0.46	3.44	0.74	0.69	0.65	0.64
GGH	-0.79	0.00019	64	2.13	0.85	3.15	0.73	0.78	0.64	0.59
ITPR2	0.49	0.00794	68	1.55	0.51	2.08	0.73	0.64	0.69	0.64
SELL	-0.52	0.00173	95	1.55	0.39	2.70	0.72	0.66	0.65	0.64
PAGE4	0.70	0.00012	82	1.82	0.66	2.55	0.72	0.74	0.64	0.55
CFAP97	0.54	0.00599	68	1.45	0.65	2.33	0.72	0.73	0.76	0.77
APOC3	-0.55	0.00054	100	1.72	0.27	3.00	0.72	0.63	0.71	0.62
TSBP1	0.46	0.00875	72	1.13	0.47	2.26	0.71	0.72	0.63	0.76
ASTN2	-0.63	0.00164	59	0.59	0.44	2.59	0.70	0.75	0.65	0.59
CCDC93	-0.42	0.00927	99	1.57	0.29	2.08	0.69	0.68	0.71	0.61
NYAP2	0.57	0.00706	59	1.57	0.08	1.82	0.69	0.84	0.71	0.65
APOC2	-0.40	0.00991	99	1.40	0.41	2.84	0.68	0.71	0.68	0.77
MYLK	0.69	0.00004	89	1.17	0.57	3.33	0.68	0.70	0.64	0.64
MGAT4A	0.61	0.00132	59	1.24	0.47	2.46	0.68	0.65	0.68	0.76
PEPD	-0.73	0.00009	86	1.40	1.07	3.55	0.68	0.63	0.65	0.63
EIF3L	-0.61	0.00575	59	1.74	0.57	2.58	0.67	0.70	0.64	0.63
SPECC1	0.77	0.00019	61	0.75	0.83	1.98	0.67	0.64	0.72	0.76
SLX4	0.54	0.00523	68	1.80	0.23	2.80	0.67	0.79	0.70	0.69
LCP1	-0.53	0.00105	95	1.94	0.62	2.68	0.67	0.67	0.58	0.63
PRG4	-0.53	0.00090	100	1.99	0.64	3.26	0.66	0.67	0.63	0.56
UACA	0.49	0.00366	90	1.47	0.79	2.69	0.66	0.61	0.68	0.65
NPM2	0.44	0.00898	95	1.35	0.42	2.68	0.66	0.67	0.63	0.65
C9	0.49	0.00233	100	0.81	1.02	2.99	0.65	0.66	0.65	0.57
PPBP	-0.73	0.00018	72	1.38	0.25	3.01	0.65	0.61	0.61	0.61

When ranking by cross validated predictive accuracy, F2R Like Trypsin Receptor 1 (F2RL1), Integrator Complex Subunit 1 (INTS1), and Ceruloplasmin (CP) were the top 3 candidate markers. This shows that the proteins that are the top hits in a cross-sectional single cohort analysis does not necessarily yield the same top hits as when assessing diagnostic accuracy on new data *in silico*. Additionally, when looking at AUC for a cross validated diagnostic ROC curve, Neuronal Tyrosine-Phosphorylated Phosphoinositide-3-Kinase Adaptor (NYAP2) displayed the highest AUC of 84%.

From the cross validated analyses, it was discovered that a single marker would not produce an AUC above 0.82. It was therefore explored whether a panel of biomarkers would potentially be useful in predicting PD diagnosis with a higher accuracy than a single biomarker. To build an optimal panel a (multiple) Support Vector Machine Recursive Feature Elimination (mSVM-RFE) approach was used with AUC and accuracy as outcome measures. 4 iterations of a 5-fold cross validation of the SVM-RFE algorithm were run, and the output model was used on the data to generate an ROC, and AUC was calculated. The proteins were then ranked in order of average importance, and the bottom 10 % of protein features were then eliminated, and the algorithm reapplied to the remaining proteins. For the last 150 proteins, one protein was eliminated at a time. Figure 51 shows the cumulative area under curve (AUC) scores as well as prediction accuracy (ACC) for addition of more features in the SVM to discriminate PD patients from healthy controls. Approximately 10 proteins were required to discriminate PD with and AUC or ACC of 90%. About 23 proteins were needed for a 100% AUC for our dataset, although including too many variables will likely overfit the model.

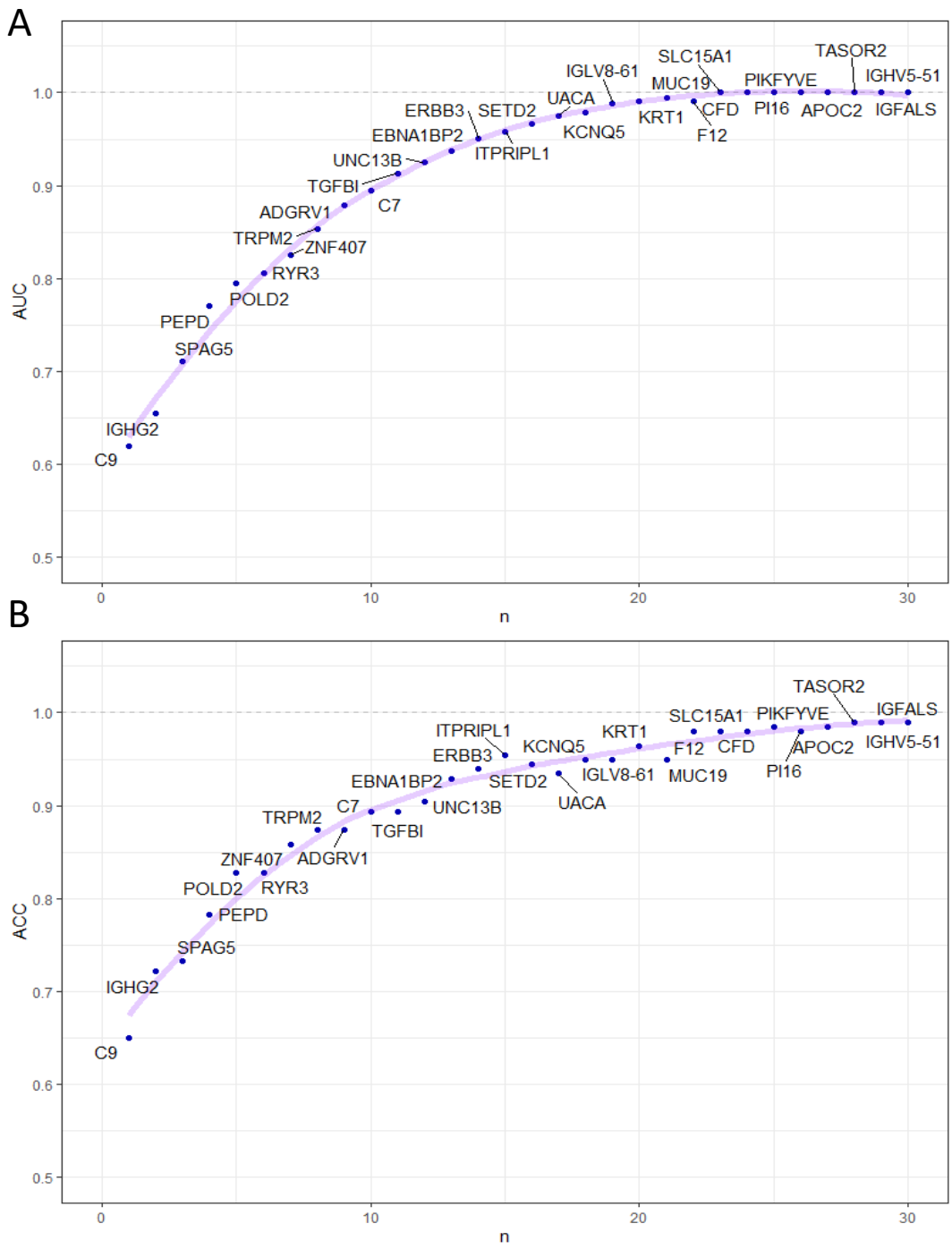

FIGURE 51 – CUMULATIVE AREA UNDER CURVE FROM A (MULTIPLE) SUPPORT VECTOR MACHINE RECURSIVE FEATURE ELIMINATION MODEL (MRFE-SVM)

A) mRFE-SVM model with increased area under curve (AUC) with increased number of proteins for differentiation of PD from healthy controls. B) mRFE-SVM model with increased diagnostic accuracy (ACC) with increased number of proteins for differentiation of PD from healthy controls.

3.1.4.2.6 BIOMARKERS FOR DISEASE SEVERTIY

Many diagnostic biomarkers that reflect disease pathology, are altered with disease severity. Therefore, identifying biomarkers that both reflect PD diagnosis as well as disease stage or disease severity, would arguably give confidence in that the protein is reflecting pathological mechanisms. Determining the best outcome for disease stage or disease severity in PD is not straight forwards. Often, UPDRS or Hoehn & Yahr (H&Y) staging is used. Particularly UPDRS part III is used to assess motor severity, and H&Y is used for symptom severity/disease staging which is also largely related to motor symptoms. Although these assessments give a crude idea of disease severity, they do not reflect all aspects of the disease progression, particularly the NMS.

Cross sectional analyses were used using robust linear models of H&Y and UPDRS III in the PD cohort versus protein quantifications. Additionally, disease duration and levodopa equivalent daily dose were separately assessed as they are often colinear with disease severity, as seen in PCA plot in figure 20. Age and sex were used as covariates. The results are presented in Table 12. The results represent the proteins that were both significantly associated with H&Y and UPDRS III and ordered by ascending H&Y p-value.

Interestingly, none of the top 30 proteins that correlated with Disease duration and LEDD correlated with H&Y and UPDRS III. This suggest that at least the top significant proteins that correlate with H&Y and UPDRS III are probably not heavily affected by disease duration and levodopa medication.

To gain confidence in which proteins were associated with disease progression/severity, at least in terms of motor progression, it was investigated which of the proteins were significantly associated with both H&Y as well as UPDRS III. The results are shown in Table 13. Serum amyloid A1 (SAA1) was the protein most significantly associated with both H&Y and UPDRS III.

Analysis of individual peptides from the proteins that correlated with both H&Y and UPDRS III was performed. Two peptides from Insulin-like growth factor-binding protein complex acid labile subunit (IGFALS), DLHFLEELQLGHNR and NLPEQVFR were associated with both H&Y ($p = 0.037$ and $p = 0.020$ respectively) and UPDRS III ($p = 0.045$ and $p = 0.021$ respectively).

TABLE 12 – CORRELATIONS WITH DISEASE SEVERITY AND MOTOR FUNCTION IN PARKINSON'S DISEASE PATIENTS

Linear model of between plasma protein levels and Hoehn & Yahr disease stage, Unified Parkinson's Disease Rating Scale (UPDRS) III motor severity, Disease Duration and Levodopa Equivalent Daily Dose (LEDD). Age and sex were used as covariates. Top 30 hits for each variable listed as Gene name of the protein, t-values and p-values listed. ordered by unadjusted p-values.

Hoehn & Yahr			UPDRS III			Disease duration			LEDD		
Gene	t-value	p-value	Gene	t-value	p-value	Gene	t-value	p-value	Gene	t-value	p-value
SAA1	4.0	0.0001	*PRR9*	4.1	0.0001	*IL16*	-4.1	0.0001	*C9*	4.1	0.0001
TRIM33	3.6	0.0006	*ZNFX1*	-3.6	0.0005	*KMT2D*	-3.9	0.0002	*DOK6*	-4.0	0.0001
ARMCX5	-3.5	0.0008	*INPP5F*	-3.6	0.0005	*CEP120*	-3.9	0.0002	*DOCK11*	4.0	0.0001
SAA2	3.2	0.0022	*CD163*	-3.3	0.0014	*KIF26B*	3.6	0.0006	*CEP120*	-3.8	0.0003
PCDHB5	3.2	0.0023	*SAA1*	3.1	0.0021	*ALOXE3*	-3.5	0.0007	*FSIP1*	3.7	0.0003
TERF2	-3.1	0.0026	*COL6A6*	3.0	0.0031	*DOK6*	-3.5	0.0008	*TRIM42*	3.7	0.0003
H2AW	-3.1	0.0030	*ALAS2*	-3.0	0.0032	*KLHL31*	-3.3	0.0012	*ATP11A*	3.6	0.0005
RFX7	-3.0	0.0030	*GPR158*	-3.0	0.0032	*AMY1A*	-3.3	0.0014	*PLK2*	-3.5	0.0007
FAM81B	3.0	0.0035	*ZC3H14*	-3.0	0.0033	*ARMC6*	-3.3	0.0016	*PALB2*	-3.4	0.0013
AHCYL1	-3.0	0.0035	*SPECC1*	-3.0	0.0036	*MYH1*	3.2	0.0018	*ADGRB3*	-3.2	0.0017
GAS2L3	-2.9	0.0043	*TIAM2*	-2.9	0.0039	*POLR2A*	3.2	0.0019	*DDX5*	-3.2	0.0017
PPP1R12A	-2.9	0.0055	*APOF*	2.8	0.0057	*ATG16L1*	3.2	0.0019	*LPA*	3.2	0.0019
COL25A1	2.8	0.0056	*CD101*	2.8	0.0058	*ADGRB3*	-3.2	0.0021	*HERC5*	-3.1	0.0027
ATP10A	2.8	0.0064	*SMARCAD1*	-2.8	0.0062	*ZNF512*	3.1	0.0025	*NDST1*	-3.0	0.0030
SNAP47	2.8	0.0065	*DENND2A*	-2.8	0.0066	*ANKRD20A8P*	-3.1	0.0027	*COMP*	3.0	0.0034
TCEA1	2.8	0.0069	*TCF20*	-2.7	0.0073	*SAMSN1*	3.1	0.0030	*SVIL*	3.0	0.0035
TSHZ3	2.8	0.0070	*SAA2*	2.7	0.0073	*ADGB*	3.0	0.0037	*CABIN1*	-3.0	0.0036
JPH3	-2.7	0.0077	*GPR179*	-2.8	0.0075	*MYO15A*	-2.9	0.0039	*IGHV4-28*	-3.0	0.0037
PLCL2	-2.7	0.0077	*RFX7*	-2.7	0.0075	*FAN1*	2.9	0.0041	*FBLN5*	2.9	0.0043
NAV1	2.7	0.0082	*SECISBP2L*	-2.7	0.0084	*CCDC66*	2.9	0.0049	*ADGB*	2.9	0.0044
CFAP65	-2.7	0.0085	*PRELID3B*	2.7	0.0085	*ARF3*	-2.9	0.0051	*CDC42BPB*	2.9	0.0047
FCN2	-2.7	0.0086	*ACBD5*	-2.7	0.0087	*MSH6*	2.9	0.0053	*COL17A1*	-2.9	0.0049
PRR9	2.7	0.0086	*VPS13D*	-2.7	0.0088	*FSIP1*	2.8	0.0056	*SERPINA2*	2.8	0.0058
EIF4ENIF1	-2.7	0.0086	*CASZ1*	-2.7	0.0090	*DDX18*	2.8	0.0059	*PDS5A*	2.8	0.0062
RAD50	-2.7	0.0087	*ZSCAN32*	-2.7	0.0091	*ESYT1*	2.8	0.0060	*IGHV3-49*	-2.8	0.0066
DHX16	2.7	0.0088	*GPRASP1*	-2.7	0.0092	*COMP*	2.8	0.0065	*ADAMTS20*	-2.8	0.0068
KMT2B	-2.7	0.0091	*ARMCX5*	-2.7	0.0093	*COL4A2*	2.7	0.0072	*NLRP14*	2.7	0.0074
RIPK1	-2.7	0.0095	*GUF1*	-2.6	0.0107	*SVIL*	2.7	0.0073	*ZNF512*	2.7	0.0076
RP1L1	2.6	0.0098	*SYNM*	-2.6	0.0112	*WDR11*	-2.8	0.0073	*THUMPD2*	2.7	0.0085
SACS	2.6	0.0104	*HBB*	-2.6	0.0114	*ESF1*	2.7	0.0078	*LRRFIP2*	2.7	0.0085

Interestingly, there was sparse overlap between PD diagnostic and severity markers. Pleckstrin Homology Like Domain Family B Member 2 (PHLDB2) was the only significant PD diagnostic protein

that was also associated with H&Y and UPDRS III, and the significance levels for all analyses were moderate. It was hypothesised this was partly due to large collinearity between age and disease severity, and proteins levels corrected for age would be overcorrected for disease severity. When age was removed as a covariate, a few more proteins were significantly associated with UPDRS III and H&Y. Fibronectin 1 (FN1), RRN3 Homolog, RNA Polymerase I Transcription Factor (RRN3), Izumo Sperm-Egg Fusion 1 (IZUMO1), CD5 Molecule Like (CD5L), Dynein Axonemal Heavy Chain 8 (DNAH8), Phosphodiesterase 6A (PDE6A) were all significant ($p<0.05$), however the significance level for all these proteins were modest for both PD diagnosis and disease severity. This suggests PD motor severity candidates are generally not overlapping with diagnostic candidates.

The analysis for proteins correlating with of H&Y and UPDRS III was also performed for males and females separately. Again a robust linear model was analysed with age and disease duration as covariates, with the most significant results for both H&Y and UPDRS III presented in Table 14. Males and females had different proteins correlating with motor severity. This could be attributed to statistical reasons such as more severe motor symptoms in the female cohort, the larger male cohort, or false positives due to small a sample size relative to number of proteins analysed. The difference could however also be attributed to biological differences. As was seen in the sex-specific analysis for PD diagnostic biomarkers, proteomic changes in PD seem to differ between males and females.

TABLE 13 – OVERLAPPING CORRELATIONS WITH DISEASE SEVERITY AND MOTOR FUNCTION IN PARKINSON'S DISEASE PATIENTS

Linear model of between plasma protein levels and Hoehn & Yahr disease stage, UPDRS III motor severity, Disease Duration and Levodopa Equivalent Daily Dose. Age and sex were used as covariates. Overlapping significant hits for H&Y and UPDRS III listed as Gene name of the protein, t-values and p-values listed. ordered by unadjusted p-values.

			Hoehn & Yahr		*UPDRS III*		*Disease duration*		*LEDD*	
Gene	*Ave Exprs*	*% quant*	*t-value*	*p-value*	*t-value*	*p-value*	*t-value*	*p-value*	*t-value*	*p-value*
SAA1	2.58	95	4.01	0.00010	3.14	0.00214	0.30	0.76179	1.26	0.21162
ARMCX5	2.62	77	-3.45	0.00081	-2.65	0.00933	-0.31	0.75894	0.54	0.59200
SAA2	2.82	73	3.15	0.00218	2.74	0.00735	0.25	0.80393	0.61	0.54411
TERF2	2.88	96	-3.08	0.00255	-2.40	0.01810	0.20	0.84429	0.00	0.99814
H2AW	2.40	56	-3.08	0.00296	-2.45	0.01681	0.12	0.90668	-0.78	0.43621
RFX7	2.52	100	-3.02	0.00305	-2.72	0.00752	-0.50	0.61626	-1.45	0.14844
TSHZ3	2.02	76	2.76	0.00696	2.27	0.02539	0.81	0.41728	0.73	0.46769
CFAP65	2.03	53	-2.71	0.00845	-2.46	0.01665	1.83	0.07099	1.56	0.12409
PRR9	2.99	82	2.68	0.00864	4.08	0.00009	0.84	0.40190	0.95	0.34557
RAD50	1.87	81	-2.67	0.00871	-2.42	0.01706	2.02	0.04606	-0.05	0.96101
ARMCX1	1.91	54	-2.49	0.01529	-2.12	0.03776	0.04	0.96847	-0.54	0.59175
RACGAP1	2.83	60	-2.46	0.01620	-2.48	0.01545	-1.15	0.25222	-1.16	0.24840
KANSL1	2.31	65	2.45	0.01643	2.07	0.04163	0.76	0.45052	1.92	0.05849
ZNF419	2.00	63	-2.36	0.02051	-2.15	0.03430	1.09	0.27710	0.37	0.71524
EXOC4	2.52	69	-2.35	0.02119	-2.18	0.03203	-0.39	0.69678	-1.10	0.27415
PHLDB2	2.16	81	-2.31	0.02293	-2.05	0.04256	-0.44	0.65962	-2.17	0.03197
MAP3K13	2.39	54	2.30	0.02453	2.01	0.04885	0.11	0.91446	0.01	0.99061
IGFALS	2.87	100	-2.24	0.02698	-1.99	0.04912	0.51	0.60917	0.47	0.63881
AFF2	2.52	73	-2.23	0.02805	-2.54	0.01259	-0.55	0.58497	-1.80	0.07468
NACA	2.48	100	-2.21	0.02921	-2.39	0.01837	-0.75	0.45168	-1.50	0.13608
SHISA7	2.48	59	-2.21	0.03007	-2.25	0.02707	-2.35	0.02108	-2.61	0.01097
COL6A6	2.24	73	2.20	0.03052	3.03	0.00312	-0.02	0.98426	0.85	0.39830
INPP5F	2.16	69	-2.17	0.03269	-3.60	0.00053	2.36	0.02027	1.18	0.24100
IBTK	2.34	53	2.16	0.03450	2.44	0.01729	-1.44	0.15306	-0.45	0.65067
CCDC33	2.77	64	2.12	0.03738	2.21	0.02966	0.38	0.70732	-0.06	0.95260
ALAS2	2.23	81	-2.11	0.03753	-3.02	0.00318	-0.72	0.47170	-2.20	0.02979
FRMPD3	2.12	66	-2.11	0.03759	-2.49	0.01491	-0.22	0.82728	-0.29	0.77005
GPR158	2.26	72	-2.05	0.04277	-3.03	0.00322	0.35	0.73028	-0.84	0.40238
SHPRH	2.43	70	2.04	0.04389	2.33	0.02183	0.57	0.56845	2.26	0.02633
PROS1	3.58	100	-1.99	0.04845	-2.24	0.02690	0.30	0.76389	-0.35	0.72757

TABLE 14 – OVERLAPPING CORRELATIONS WITH DISEASE SEVERITY AND MOTOR FUNCTION IN PARKINSON'S DISEASE PATIENTS – SEX-SPECIFIC ANALYSIS

Linear model of between plasma protein levels and Hoehn & Yahr disease stage, UPDRS III motor severity. Age and disease duration were used as covariates. Overlapping significant hits for H&Y and UPDRS III listed as Gene name of the protein and p-values listed, separately analysed for male and female PD, ordered by unadjusted p-values for Hoehn & Yahr.

Female PD	*Hoehn & Yahr*	*UPDRS III*	*Male PD*	*Hoehn & Yahr*	*UPDRS III*
Gene	*p-value*	*p-value*	*Gene*	*p-value*	*p-value*
DNAI4	0.000008	0.000830	TRIM33	0.000133	0.009805
ARAP2	0.000393	0.007648	GPR158	0.000717	0.000014
COL25A1	0.000620	0.007913	SEMA4C	0.001439	0.017153
RUFY2	0.001013	0.010155	SAA2	0.001638	0.014037
ARMCX1	0.001023	0.030247	SAA1	0.002112	0.001837
DYNC2H1	0.001349	0.000048	AHCYL1	0.003383	0.015338
RIPK1	0.002361	0.010904	BAZ2B	0.003646	0.025402
EIF2A	0.003114	0.011615	TSPAN14	0.003752	0.022444
ISOC1	0.003122	0.003677	ITGA6	0.004611	0.001273
CELF5	0.003357	0.001734	SNCAIP	0.005011	0.031194
MAGEB6	0.004017	0.009575	MACO1	0.005167	0.002553
MAP7D3	0.005271	0.000106	EPHA2	0.005326	0.028322
TUT4	0.005788	0.012292	APOC2	0.005349	0.004705
IQCC	0.006672	0.011406	ATXN7L3	0.005559	0.025156
ABCG2	0.007520	0.026701	PRELID3B	0.006315	0.003341
VASN	0.009005	0.004551	ARMCX5	0.006691	0.025530
EFCAB5	0.010092	0.029518	LONP2	0.008223	0.020386
ZNF419	0.011884	0.002252	TERF2	0.008319	0.029605
RAB11FIP5	0.012502	0.032045	PRR9	0.00857	0.002181
ABL2	0.013820	0.027389	KANSL1	0.008752	0.002707
FRY	0.014813	0.010676	IHO1	0.009827	0.006013
DSCAML1	0.015610	0.024192	SLC15A1	0.013177	0.027934
ACACA	0.016680	0.022934	EVC	0.013542	0.018258
NLRP10	0.021504	0.001680	MAP3K13	0.016794	0.011118
CHMP1A	0.021885	0.018143	CDK9	0.019543	0.001618
AMY1A	0.022483	0.011698	PROS1	0.019709	0.006760
ZNFX1	0.023171	0.003084	COL6A5	0.022837	0.022634
GCKR	0.025084	0.000565	MCF2L2	0.022987	0.022574
SLC22A11	0.027154	0.021080	TCF20	0.029486	0.000490
C2CD6	0.030620	0.033237	MROH2A	0.029953	0.001709

3.1.4.2.7 PATHWAY ANALYSIS

Finally, macrolevel analysis of the significantly differentially expressed proteins for PD diagnosis were performed. A protein-protein interaction plot was constructed using all significantly differentially expressed proteins with a Cohen's d effect size >0.5 (Figure 52). Furthermore, all 187 significant proteins were used in a pathway analysis (Table 15). After adjusting for multiple comparison, the KEGG pathway "Complement and coagulation cascades" was still highly significantly implicated in PD ($p<0.00001$). These findings warrant further exploration of the role of the peripheral complement and coagulation systems in PD.

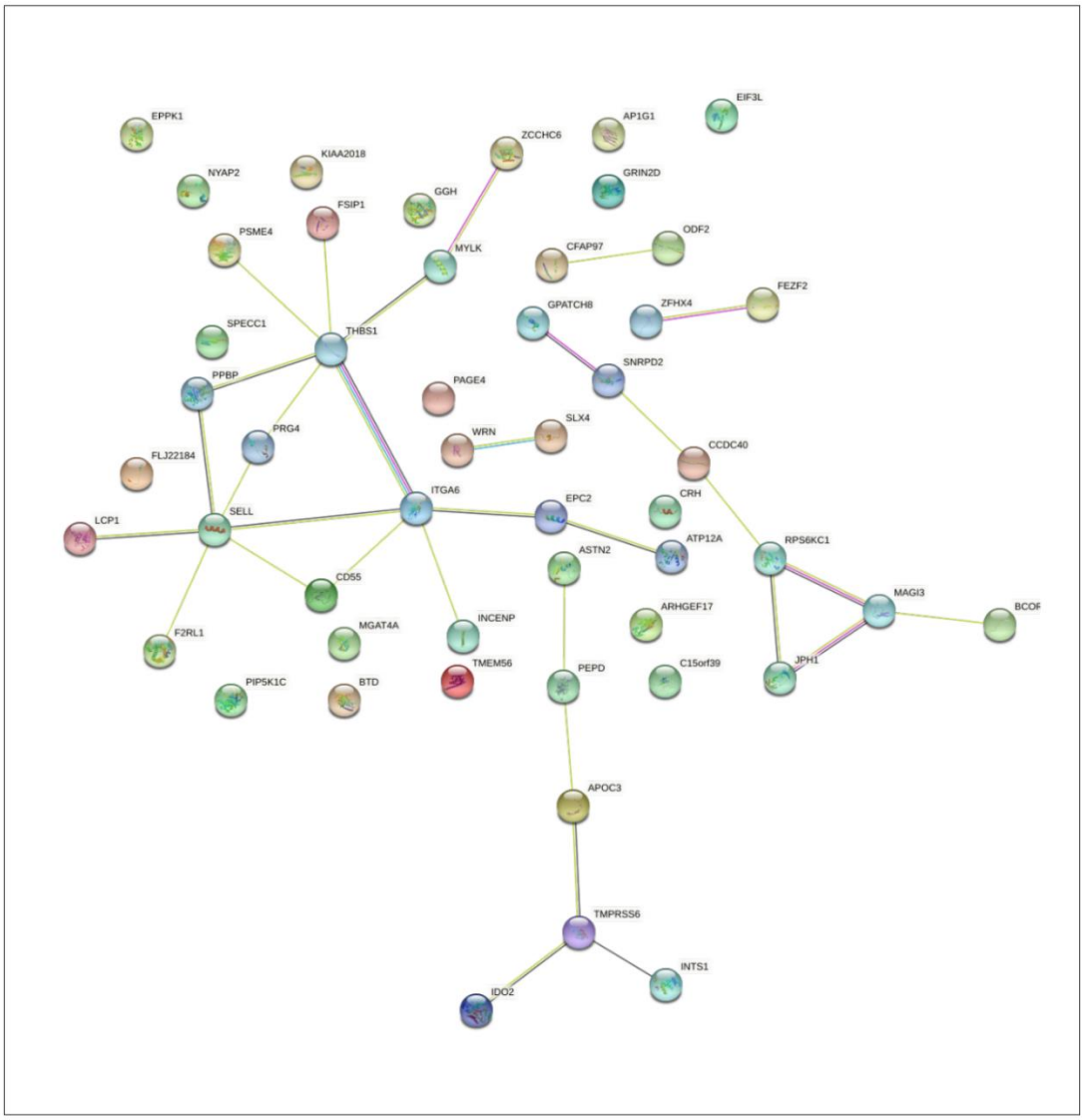

FIGURE 52 – PROTEIN-PROTEIN INTERACTIONS OF MOST SIGNIFICANT PROTEINS FOR PARKINSON'S DISEASE

Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) protein-protein interaction network of significantly changed proteins in Parkinson's disease versus healthy controls (unadjusted $p<0.05$), adjusted for age and sex, with a Cohen's d effect size >0.5.

TABLE 15 – PATHWAY ANALYSIS OF ALL SIGNIFICANTLY DIFFERENTIALLY EXPRESSED PROTEINS IN PARKINSON'S DISEASE (PD)

Implicated pathways of all proteins that were significantly (p<0.05) differentially expressed in Parkinson's disease versus Healthy controls when adjusted for Age and Sex (n=187), based on KEGG (Kyoto Encyclopaedia of Genes and Genomes). Columns represent KEGG pathway, number of differentially expressed proteins overlapping with the pathway, p-values that the pathway is implicated in PD, adjusted p-values, and gene names of the differentially expressed proteins in the pathway.

KEGG pathway	*Overlap*	*p-value*	*Adj. p-value*	*Gene name of proteins*
Complement and coagulation cascades	9/79	**0.00000005**	**0.00000707**	*CFD; C7; SERPIND1; C9; F13A1; F13B; MASP1; CFB; CD55*
ECM-receptor interaction	5/82	**0.00103266**	0.06918835	*COMP; COL6A3; ITGA6; THBS1; CD44*
Homologous recombination	3/41	**0.00660522**	0.29503321	*EME1; POLD2; TOPBP1*
Focal adhesion	6/199	**0.01118130**	0.37457342	*COMP; COL6A3; ITGA6; PIP5K1C; THBS1; MYLK*
Phosphatidylinositol signaling system	4/99	**0.01399301**	0.37501274	*PIKFYVE; MTMR8; ITPR2; PIP5K1C*
RNA transport	5/165	**0.01957726**	0.43722544	*NDC1; NUP210L; NUP133; TPR; THOC2*
Staphylococcus aureus infection	3/68	**0.02586736**	0.45370637	*CFD; MASP1; CFB*
Calcium signaling pathway	5/188	**0.03199559**	0.45370637	*ITPR2; ATP2A1; RYR3; MYLK; GRIN2D*
Inositol phosphate metabolism	3/74	**0.03213788**	0.45370637	*PIKFYVE; MTMR8; PIP5K1C*
Mucin type O-glycan biosynthesis	2/31	**0.03385868**	0.45370637	*GALNT5; GCNT3*
Prion diseases	2/35	**0.04228978**	0.51516646	*C7; C9*
African trypanosomiasis	2/37	**0.04676876**	0.52225117	*F2RL1; IDO2*
Hematopoietic cell lineage	3/97	0.06270793	0.62724637	*ITGA6; CD44; CD55*
Malaria	2/49	0.07684995	0.62724637	*COMP; THBS1*
Alzheimer disease	4/171	0.07691541	0.62724637	*ITPR2; ATP2A1; RYR3; GRIN2D*
Cholesterol metabolism	2/50	0.07957603	0.62724637	*APOC2; APOC3*
N-Glycan biosynthesis	2/50	0.07957603	0.62724637	*MGAT4A; MAN1A1*
Fanconi anemia pathway	2/54	0.09077091	0.67573902	*SLX4; EME1*

To summarise, statistical analysis of the proteomic data revealed several candidate markers for PD with moderate predictive performance. Firstly, it was explored whether the entire plasma proteome would generate any separation or clustering of the data according to diagnosis. There was no apparent separation in the data on proteome level, hence individual differentially expressed plasma proteins were analysed. A p-value histogram revealed at least 50 proteins were truly altered in PD. The data was analysed with and without age and sex as covariates, and with and without an effect size cut-off. It was observed age and sex had little impact on top significant candidates for PD. Sorting by effect size reordered the rank of the most differentially expressed candidates, as it was independent of sample size. The proteins ranked by effect size are arguably better biomarker candidates as they would be the most resistant to within-group variation. Moreover, sex-specific analysis revealed different sets of altered plasma proteins in male and female PD patients.

Next, the predictive diagnostic performance of the candidate biomarkers was assessed, as that would reflect the potential clinical utility of the biomarkers. Hence, the dataset was split multiple times to train a predictive model to test on a subset of data. The highest diagnostic accuracy achieved was 76%, and it was noticed the proteins with the highest predictive performance, were not necessarily the most significant ones, nor the ones with biggest effect size. Furthermore, it provided insight in how well the candidate markers would perform on new data. It was also attempted to build a predictive biomarker panel using several proteins, and it was found approximately 10 proteins were necessary to reach a 90% diagnostic accuracy.

It was hypothesised that diagnostic biomarkers that reflect disease pathology, would also be associated with disease progression and severity. Several proteins were found that correlated with motor severity, that were not affected by covariates such as age, medication, and disease duration. However, very few of these overlapped with the diagnostic markers. There are a few possible explanations for this. It is possible the diagnostic biomarkers reflect function defects and risk factors that lead to PD, whereas progression markers are responsive and reflect cell death and plaques.

Finally, pathways and protein interactions implicated with the top biomarker candidates were analysed. The complement and coagulation pathways were the most significant ones, and several of the biomarkers had known protein-protein interactions, particularly several structural and adhesion molecules.

The top candidate biomarkers for both diagnosis and disease severity were selected for the verification study.

3.1.4.3 PD cognition biomarkers

3.1.4.3.1 DIMENTIONALITY REDUCTION

Besides identifying diagnostic PD markers, one of the most important aims related to PD endophenotype biomarkers was which plasma proteins would reflect cognitive impairment and cognitive decline in PD. Just as with the diagnostic markers, initially data dimensionality reduction methods were applied to get an overview of the cognitive data. Few patients in the cohort had an actual PDD diagnosis set by a physician, but all of them had MoCA scores available. PDND, PDMCI and PDD thresholds were set at MoCA ≥26, 21-25, and ≤20 respectively, as these were recommended thresholds from the literature for PD patients that develop MCI and dementia [175]. It was also made sure there was no other documented comorbidity (infection, Alzheimer's disease, stroke etc) that could explain the cognitive score.

PCA and t-SNE plots showed little discrimination between groups on proteome level (Figure 53 and Figure 54).

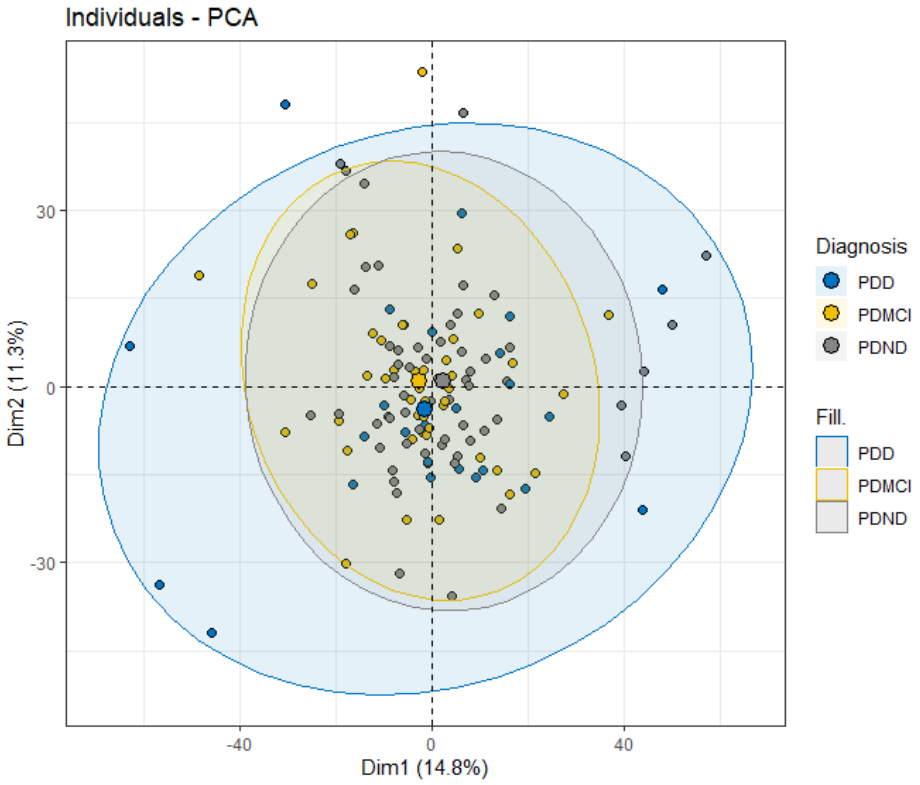

FIGURE 53 – PRINCIPAL COMPONENT ANALYSIS (PCA) PLOT OF PROTEOMIC DATA

PCA plot of mass spectrometric data of Parkinson's disease (PD) patients. Patients classified as PD with no dementia (PDND), PD with mild cognitive impairment (PDMCI), and PD with dementia (PDD) based on conventional reference levels of Montreal Cognitive Assessment (MoCA). Centre point enlarged; ellipse shows 95% confidence interval. Principal components 1 and 2 shown. Variables are centred and scaled.

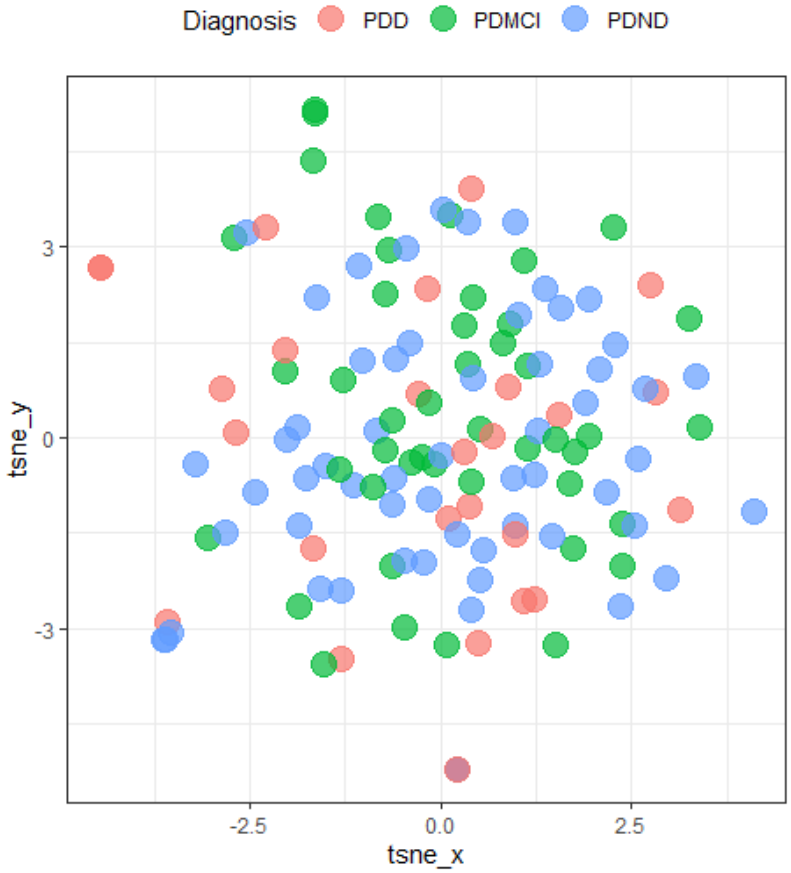

FIGURE 54 – T-DISTRIBUTED STOCHASTIC NEIGHBOR EMBEDDING (TSNE) OF PROTEOMIC DATA

t-SNE plot of mass spectrometric data of Parkinson's disease (PD) patients. Patients classified as PD with no dementia (PDND), PD with mild cognitive impairment (PDMCI), and PD with dementia (PDD) based on conventional reference levels of Montreal Cognitive Assessment (MoCA).

3.1.4.3.2 CORRELATION WITH COGNITIVE PERFORMANCE

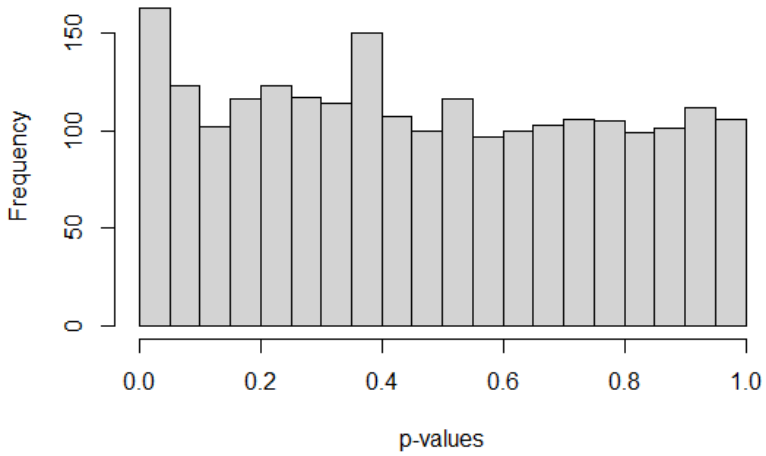

FIGURE 55 – P-VALUE HISTOGRAM FROM LINEAR MODEL CORRELATION BETWEEN MOCA SCORES AND PROTEIN LEVELS

Next a robust linear model was fit with MoCA as the outcome variable, and protein quantification, age, sex, and disease duration as explanatory variables in the limma package. The top significant proteins are listed in Table 16.

A p-value histogram was plotted to visualise whether there were true correlations between protein levels and MoCA score (Figure 55). The higher-than-average bar representing p-values between 0-0.05 on the x-axis suggest that there is a true correlation with MoCA for up to 50 proteins in the PD patients, and that the null hypothesis should be rejected for up to 50 protein IDs.

Unlike in the group analysis of PD vs healthy controls, the PD patients across all cognitive scores were not well age and sex matched across all cognitive levels. People with worse cognitive scores tended to be older and more often of male sex, which is generally the case in PD patients. Moreover, cognitive performance is well documented to decline over time with the disease. Hence the statistics for age, sex and disease duration were evaluated, and proteins that had strong correlation with any of these covariates were highlighted in Table 17. These were considered not robust biomarker candidates as they demographic variables could affect the protein levels too much relative to the cognitive performance.

TABLE 16 – TOP 30 PROTEINS FOR ASSOCIATED WITH LEVEL OF COGNITIVE FUNCTION IN PARKINSON'S DISEASE

Top results from robust linear model of proteins in Parkinson's Disease versus Montreal Cognitive Assessment (MoCA) score, with age, sex, and disease duration as covariates, ranked by p-value. Columns show Protein name, Gene name, average expression level of protein (AveExpr), t-statistic, p-value, Benjamini-Hochberg (BH) adjusted p-value, and percentage of samples where the protein was quantified.

Protein	*Gene*	*AveExprs*	*t-value*	*p-value*	*adj p-value*	*% quantified*
Serum amyloid A-1 protein	SAA1	2.58	-3.93	0.00014	0.254	95
IQ domain-containing protein C	*IQCC*	1.97	3.85	0.00023	0.254	68
Microtubule-associated protein 10	*MAP10*	2.50	-3.61	0.00050	0.379	68
GAS2-like protein 3	*GAS2L3*	2.68	3.45	0.00093	0.479	55
Protein PHTF2	*PHTF2*	2.31	-3.26	0.00171	0.479	54
Protein BCAP	*ODF2L*	2.56	3.21	0.00178	0.479	82
Huntingtin	*HTT*	2.23	-3.23	0.00186	0.479	55
Tumor protein 63	*TP63*	3.25	3.16	0.00195	0.479	95
Slit homolog 3 protein	*SLIT3*	1.89	-3.17	0.00204	0.479	73
HLA class II histocompatibility antigen gamma chain	*CD74*	2.70	-3.16	0.00233	0.479	55
Golgi resident protein GCP60	*ACBD3*	2.25	3.10	0.00251	0.479	82
Transthyretin	*TTR*	3.58	3.08	0.00254	0.479	100
Vitamin K-dependent protein S	*PROS1*	3.58	3.05	0.00279	0.485	100
Biotinidase	*BTD*	2.56	3.03	0.00329	0.505	58
Collagen alpha-1(XII) chain	*COL12A1*	2.33	-3.00	0.00353	0.505	68
Ras GTPase-activating protein-binding protein 2	*G3BP2*	2.51	2.97	0.00369	0.505	82
Nebulette	*NEBL*	2.36	2.93	0.00411	0.505	82
Glutathione peroxidase 3	*GPX3*	2.48	2.94	0.00419	0.505	68
Immunoglobulin heavy variable 1-3	*IGHV1-3*	2.25	-2.90	0.00451	0.505	91
Acetyl-CoA carboxylase 2	*ACACB*	2.24	2.90	0.00475	0.505	64
AP2-interacting clathrin-endocytosis protein	*KIAA1107*	2.27	2.89	0.00478	0.505	72
Zinc finger protein 419	*ZNF419*	2.00	2.89	0.00491	0.505	63
Pecanex-like protein 1	*PCNX1*	2.33	-2.85	0.00565	0.517	54
Alcohol dehydrogenase 6	*ADH6*	2.88	2.80	0.00615	0.517	78
FRAS1-related extracellular matrix protein 1	*FREM1*	2.34	-2.81	0.00624	0.517	59
Zinc finger and SCAN domain-containing protein 32	*ZSCAN32*	2.22	2.78	0.00665	0.517	63
Nascent polypeptide-associated complex subunit alpha, muscle-specific form	*NACA*	2.48	2.76	0.00667	0.517	100
Zinc finger protein castor homolog 1	*CASZ1*	2.36	2.77	0.00693	0.517	64
Junctophilin-3	*JPH3*	2.75	2.75	0.00703	0.517	77
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1	*SMARCAD1*	2.31	2.77	0.00721	0.517	55

TABLE 17 – TOP 30 PROTEINS FOR ASSOCIATED WITH LEVEL OF COGNITIVE FUNCTION IN PARKINSON'S DISEASE – INTERFERING COVARIATES

Top results from robust linear model of proteins in Parkinson's Disease versus Montreal Cognitive Assessment (MoCA) score, with age and sex as covariates, ranked by p-value. Columns show, Gene name, t-statistic, and p-value. T-values and p-values are also displayed for Age, Sex and Disease duration. Proteins where covariates are associated (p<0.05) with the protein are highlighted.

	MoCA		Age		Sex		Disease duration	
Gene	t-value	p-value	t-value	p-value	t-value	p-value	t-value	p-value
SAA1	-3.926	0.00014	0.305	0.76113	0.350	0.72662	0.623	0.53458
IQCC	3.847	0.00023	0.147	0.88351	-0.463	0.64435	1.675	0.09750
MAP10	-3.612	0.00050	-1.370	0.17399	0.497	0.62066	-0.444	0.65809
GAS2L3	3.453	0.00093	0.669	0.50566	-0.408	0.68447	-0.249	0.80418
PHTF2	-3.262	0.00171	-0.482	0.63152	2.118	0.03766	-0.738	0.46273
ODF2L	3.206	0.00178	0.988	0.32535	-0.448	0.65502	-1.477	0.14255
HTT	-3.232	0.00186	-4.543	0.00002	-0.341	0.73436	0.739	0.46252
TP63	3.164	0.00195	-0.624	0.53372	-2.527	0.01274	-0.455	0.65002
SLIT3	-3.170	0.00204	0.047	0.96287	1.120	0.26532	-1.742	0.08466
CD74	-3.156	0.00233	-3.319	0.00141	-1.866	0.06611	0.134	0.89372
ACBD3	3.096	0.00251	1.578	0.11749	-0.258	0.79686	0.010	0.99219
TTR	3.077	0.00254	0.180	0.85761	-2.102	0.03746	-0.891	0.37439
PROS1	3.048	0.00279	0.728	0.46779	-0.072	0.94288	0.106	0.91572
BTD	3.034	0.00329	-0.574	0.56773	-0.343	0.73252	1.673	0.09838
COL12A1	-2.998	0.00353	-0.394	0.69489	1.090	0.27857	-0.502	0.61664
G3BP2	2.968	0.00369	0.613	0.54120	0.653	0.51517	1.350	0.17985
NEBL	2.933	0.00411	-0.529	0.59815	-0.652	0.51570	-0.935	0.35185
GPX3	2.940	0.00419	-2.879	0.00499	-1.976	0.05129	0.273	0.78525
IGHV1-3	-2.896	0.00451	1.120	0.26486	-0.766	0.44524	-0.313	0.75510
ACACB	2.901	0.00475	2.073	0.04124	0.950	0.34476	-1.474	0.14434
KIAA1107	2.889	0.00478	1.908	0.05936	-0.575	0.56688	-0.520	0.60425
ZNF419	2.890	0.00491	0.472	0.63789	-0.571	0.56975	0.963	0.33815
PCNX1	-2.855	0.00565	-2.153	0.03469	1.462	0.14831	-1.887	0.06332
ADH6	2.798	0.00615	1.657	0.10051	1.299	0.19677	0.709	0.47983
FREM1	-2.811	0.00624	-1.335	0.18589	-0.388	0.69885	0.677	0.50026
ZSCAN32	2.784	0.00665	-1.690	0.09483	0.639	0.52478	0.024	0.98075
NACA	2.757	0.00667	0.854	0.39444	-0.907	0.36585	-0.789	0.43151
CASZ1	2.769	0.00693	0.457	0.64913	2.331	0.02215	1.362	0.17674
JPH3	2.752	0.00703	-1.102	0.27316	-0.978	0.33058	0.534	0.59478

The top 4 significant proteins, Serum amyloid A-1 protein, IQ domain-containing protein C, Microtubule-associated protein 10, and GAS2-like protein 3 were plotted as scatter plots in Figure 56. As can be seen from the plot, the correlations appear strong, but the cognitive data is skewed with fewer individuals with lower MoCA scores.

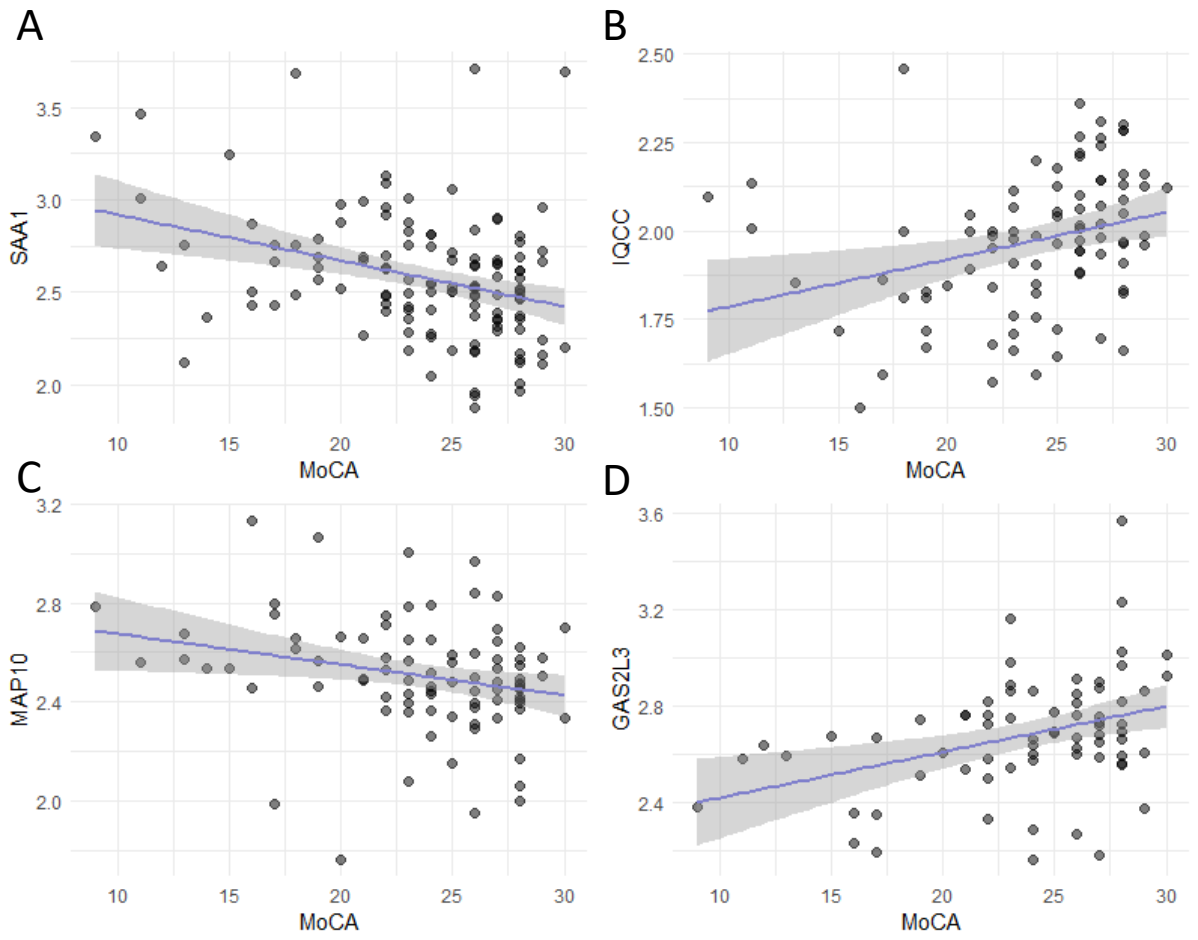

FIGURE 56 – SCATTERPLOTS OF THE TOP 4 PROTEINS CORRELATING WITH MONTREAL COGNITIVE ASSESSMENT (MOCA) SCORE IN PATIENTS WITH PARKINSON'S DISEASE

A) Plasma Serum amyloid A-1 protein (SAA1) levels versus MoCA score. B) Plasma IQ domain protein C (IQCC) levels versus MoCA score. C) Plasma Microtubule-associated protein 10 (MAP10) levels versus MoCA score. D) Plasma and GAS2-like protein 3 (GAS2L3) levels versus MoCA score.

Owing to the large covariance between motor symptom severity and cognitive impairment, several of the proteins associated with cognitive performance were also associated with motor severity. These included Nascent polypeptide associated complex subunit alpha (NACA), Protein S (PROS1), Proline

rich 9 (PRR9), Serum Amyloid A1 (SAA1), and Zinc Finger Protein 419 (ZNF419). Interestingly, the most significant protein for both disease severity and cognitive severity was SAA1.

3.1.4.3.3 PEPTIDE ANALYSIS

Next, individual peptides from the proteins that were significantly correlated with MoCA were analysed. Again, age, sex and disease duration were added as covariates, and the statistically significant peptides are presented in Table 18. 33 peptides were correlating with MoCA scores in PD patients. The strongest correlating peptides were the EDRVIFKEMK peptide from Spectrin Repeat Containing Nuclear Envelope Protein 1 (SYNE1), the SPSQADINK peptide from Apolipoprotein B (APOB), and the ENLLDILTEPERKPDPK peptide from Cytoskeleton associated protein 2 like (CKAP2L). Similar to the analysis of diagnostic PD biomarkers, many of the peptides were only quantified in a minority of the samples. Furthermore, similar to the peptide analysis of diagnostic biomarkers, analysis of peptides from all detected proteins led to a heavy bias towards the most abundant plasma protein peptides.

TABLE 18 –PEPTIDES FROM PROTEINS IN PARKINSON'S DISEASE ASSOCIATED WITH COGNITIVE IMPAIRMENT

Analysis of individual peptides from PD plasma proteins associated with cognitive impairment. Significant results from robust linear model of peptide levels associated with Montreal Cognitive Assessment (MoCA) score in Parkinson's Disease with age, sex and disease duration as covariates. Columns show Gene name, Peptide sequence, uncorrected p-value, and percentage of samples where the peptides were quantified.

Gene	*Peptide sequence*	*p-value*	*% Quantified*
SYNE1	EDRVIFKEMK	0.001	20.0
APOB	SPSQADINK	0.003	16.9
CKAP2L	ENLLDILTEPERKPDPK	0.004	21.5
RARS2	TTKELK	0.005	40.0
APOA4	LGEVNTYAGDLQK	0.005	26.9
C4BPA	KPELVNGR	0.005	11.5
C4BPA	CEWETPEGCEQVLTGK	0.007	84.6
C2CD5	VIRLSSLNLTNQALNK	0.007	10.8
C4BPA	QSTLDKEL	0.008	70.8
TTN	TISGEIDVNVIARPSAPK	0.010	11.5
VPS13A	DGSASPAVTK	0.013	40.0
C5	DSLDQLVGGVPVTLNAQTIDVNQETSDLDPSK	0.014	14.6
APOB	SEYQADYESLR	0.016	33.1
APOB	LQDFSDQLSDYYEK	0.016	10.8
NACA	KTPAIPTPK	0.017	10.0
IGFALS	TFTPQPPGLER	0.017	27.7
APOB	IGVELTGR	0.018	100.0
APOA4	LGPHAGDVEGHLSFLEK	0.019	16.2
APOB	VAWHYDEEK	0.022	16.2
TTN	AKSVDVTEKDPMTLECVVAGTPELK	0.025	23.1
SOX30	LTKVPLTPVPTK	0.027	13.8
IGFALS	NLPEQVFR	0.028	14.6
CMYA5	ESELSKGGSVDITKETVK	0.028	6.9
MAP3K13	VKTQMSLGKLCVEER	0.032	25.4
IGFALS	SFEGLGQLEVLTLDHNQLQEVK	0.037	20.0
APOB	FSHVEK	0.042	30.0
APOB	QGFFPDSVNK	0.043	55.4
APOA4	EAVEHLQK	0.044	75.4
CFAP44	WELMMKTK	0.046	51.5
APOA4	IDQNVEELKGR	0.047	17.7
IGFALS	DLHFLEELQLGHNR	0.047	13.8
APOB	EVYGFNPEGK	0.048	85.4
APOB	NYQLYK	0.049	85.4

3.1.4.3.4 SEX-SPECIFIC ANALYSIS

As with the analysis of the diagnostic biomarkers, sex-specific analysis was conducted for the association between MoCA and protein levels (Table 19). Generally, a larger number of proteins significantly correlating with cognitive score was seen in the male PD cohort. This was not surprising given the larger number of male PD patients in the cohort, together with a greater spread in MoCA scores in the male patient cohort. The top three most significant proteins for females and males respectively are shown in Figure 57. It was noted that some of the proteins associated MoCA in the full PD cohort were more pronounces in the female cohort (such as IQCC), whereas others (such as MAP10 and SAA1) were more pronounced in the male cohort.

TABLE 19 – TOP 30 PROTEINS ASSOCIATED WITH LEVEL OF COGNITIVE FUNCTION IN PARKINSON'S DISEASE – SEX-SPECIFIC ANALYSIS

Top results from robust linear model of proteins in Parkinson's Disease versus Montreal Cognitive Assessment (MoCA) score, with age and disease duration as covariates, ranked by p-value. Females and males analysed separately. Columns show Gene name, p-value, and number of samples quantified.

Females			Males		
Gene	p-value	n quantified	Gene	p-value	n quantified
IQCC	0.00028	38	COL6A5	0.00009	59
CLIP4	0.00029	40	TP63	0.00017	69
DPP3	0.00037	32	CCT4	0.00019	46
ARHGAP45	0.00055	50	ODF2L	0.00027	60
ZRANB3	0.00075	31	FAM111B	0.00031	40
PDZD2	0.00088	36	IFT81	0.00033	61
DNAJC13	0.00104	37	ACOT12	0.00041	56
MROH2B	0.00108	38	MAP10	0.00051	50
INPP4B	0.00126	56	BTD	0.00064	45
SLU7	0.00128	30	PHTF2	0.00068	39
ATG16L1	0.00148	44	GLCCI1	0.00091	69
MED1	0.00207	48	TEX9	0.00093	43
NEK7	0.00207	40	RYBP	0.00100	54
PRPF31	0.00218	32	BRIP1	0.00100	39
ARAP2	0.00276	32	SOX30	0.00115	67
RAP1GDS1	0.00288	37	CD74	0.00115	39
PICALM	0.00313	44	ADH6	0.00127	57
FLT4	0.00338	35	COL12A1	0.00130	51
RUFY2	0.00340	31	SMC3	0.00155	46
NIPA1	0.00352	40	SEMA4C	0.00166	41
F8	0.00525	41	TTR	0.00170	72
SUSD1	0.00535	46	TAS1R3	0.00191	72
ANKRD45	0.00592	41	FREM3	0.00191	45
EML4	0.00655	36	COLEC12	0.00196	49
NRCAM	0.00725	39	SAA1	0.00199	69
XRRA1	0.00739	56	GPX3	0.00208	49
AHR	0.00742	45	IGHV1-3	0.00233	66
YEATS2	0.00777	33	ACBD3	0.00238	59
RGS22	0.00785	33	NRIP1	0.00238	48
MGME1	0.00807	30	MYH1	0.00253	40

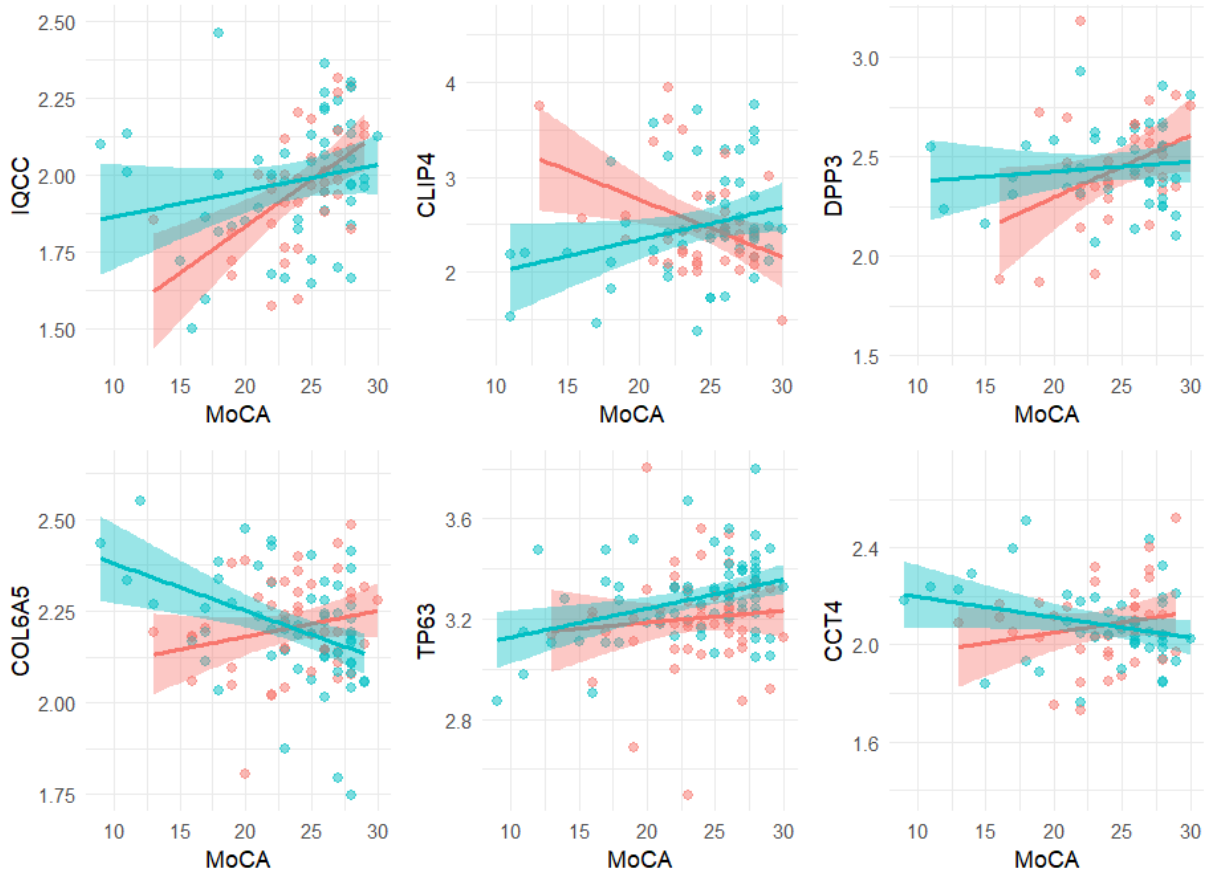

FIGURE 57 - SCATTERPLOTS OF THE TOP 3 PROTEINS FOR MALES (BLUE) AND FEMALES (RED) RESPECTIVELY CORRELATING WITH MONTREAL COGNITIVE ASSESSMENT (MOCA) SCORE IN PATIENTS WITH PARKINSON'S DISEASE

Top) The 3 most significant proteins for female PD patients that correlate with MoCA. IQ Domain-Containing Protein C (IQCC), CAP-Gly domain containing linker protein family member 4 (CLIP4), Dipeptidyl peptidase 3 (DPP3). Bottom) The 3 most significant proteins for male PD patients that correlate with MoCA. Collagen type VI alpha 5 chain (COL6A5), Tumor protein p63 (TP63), Chaperonin containing TCP1 subunit 4 (CCT4). Blue data points represent male patients, red data points represent female patients.

3.1.4.3.5 SURVIVAL ANALYSIS

Longitudinal MoCA scores were available for nearly all PD patients, but the time points were highly variable, and the plasma samples used were from different time points for each patient. Therefore, linear models were fit for cognitive decline of each patient between time of diagnosis and their last available MoCA score. MoCA<26 and MoCA<21 were set as thresholds for when they would convert to PDMCI and PDD respectively. Subsequently survival analyses with cox regression were performed for each protein to determine the hazard ratio of cognitively unimpaired PD patients converting to

MCI and dementia over time (Figure 58 and Figure 59). Details on how the model was constructed are outlined in the methods chapter (section 2.1.4.3). The survival curves and the most significant proteins for conversion to PDMCI and PDD respectively are listed in Table 20 and Table 21.

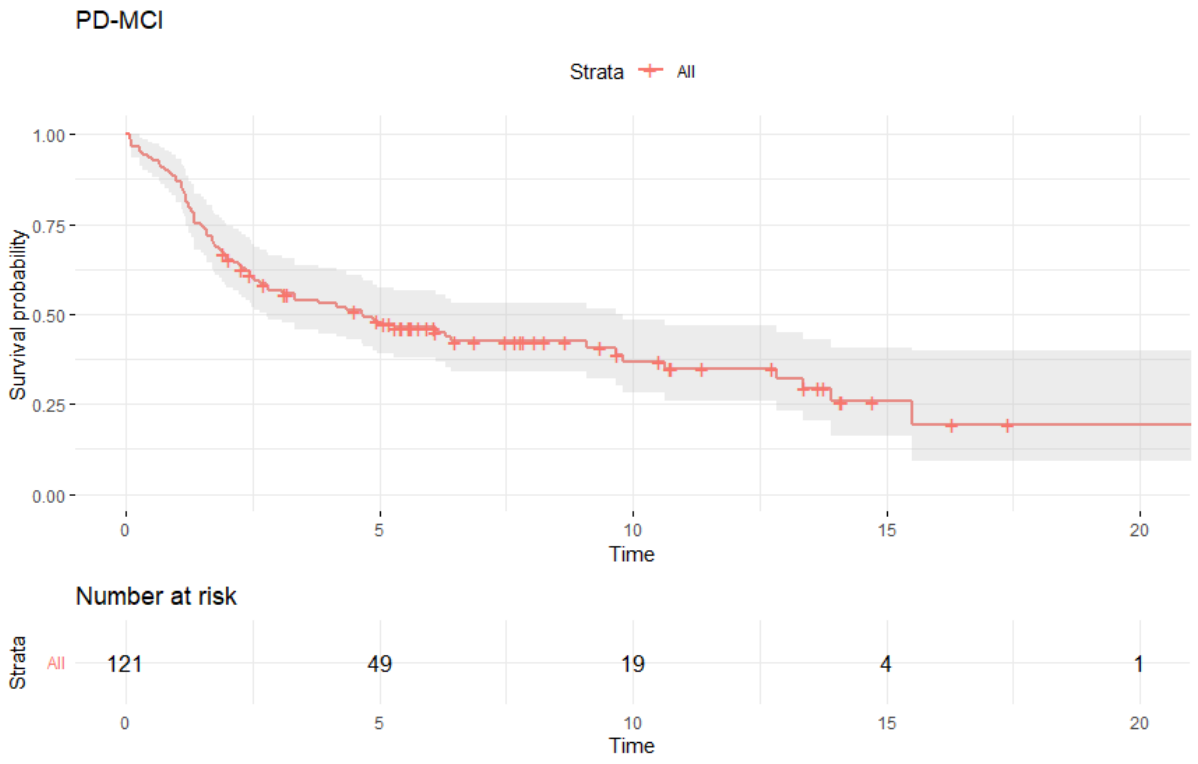

FIGURE 58 – SURVIVAL CURVE OF PARKINSON'S DISEASE (PD) PATIENTS OVER TIME THAT DEVELOP MILD COGNITIVE IMPAIRMENT (PDMCI)

Survival plot for all PD patients from the discovery study and their survival probability of declining to a Montreal Cognitive Assessment (MoCA) score of <26 over time. Time scale shows years from PD diagnosis. Trendline shows mean survival probability and 95% confidence interval.

Table 20 – Cox regression of Parkinson's Disease developing mild cognitive impairment

Cox regression of plasma protein levels in Parkinson's Disease patients versus risk of declining cognitively to Montreal Cognitive Assessment <26, with age and sex as covariates. Columns show protein gene name, Cox coefficient, Hazard ratio, standard error, z-statistic, and p-value.

		Conversion to MoCA <26 (equivalent to PD-MCI)			
Gene	*coef*	*Hazard ratio exp(coef)*	*se(coef)*	*z*	*p-value*
PROS1	-5.521	0.004	1.546	-3.571	0.00036
TERF2	-2.593	0.075	0.736	-3.523	0.00043
OBSCN	-2.813	0.060	0.810	-3.475	0.00051
PROC	-4.469	0.011	1.299	-3.441	0.00058
C2CD5	3.966	52.798	1.171	3.387	0.00071
EMILIN2	6.158	472.304	1.902	3.237	0.00121
INVS	-5.101	0.006	1.586	-3.217	0.00130
TFAP2D	-1.302	0.272	0.413	-3.155	0.00161
NWD2	2.657	14.249	0.853	3.114	0.00184
UFL1	-2.501	0.082	0.813	-3.076	0.00210
LUM	-3.918	0.020	1.285	-3.049	0.00230
SNRPD2	-2.907	0.055	0.959	-3.031	0.00244
HEATR1	2.420	11.246	0.810	2.987	0.00282
COL11A2	-3.460	0.031	1.164	-2.973	0.00295
CHD4	1.742	5.707	0.604	2.886	0.00390
LRG1	-3.186	0.041	1.107	-2.878	0.00401
ACACA	2.401	11.033	0.844	2.845	0.00444
KRT9	-0.989	0.372	0.348	-2.841	0.00450
RNF213	-2.731	0.065	0.962	-2.838	0.00454
CNN2	-2.282	0.102	0.805	-2.833	0.00461
PRKG2	-2.741	0.065	0.969	-2.827	0.00470
SLIT3	3.802	44.782	1.349	2.818	0.00484
ZNF407	-2.047	0.129	0.731	-2.802	0.00508
PDAP1	6.634	760.580	2.373	2.796	0.00518
IHO1	3.309	27.355	1.195	2.768	0.00564
DNAH1	-1.779	0.169	0.643	-2.765	0.00569
UBN2	4.554	95.000	1.654	2.754	0.00589
ARMCX1	-3.289	0.037	1.211	-2.716	0.00660
XRCC6	3.305	27.241	1.217	2.715	0.00663
ENTPD7	1.994	7.344	0.737	2.707	0.00680

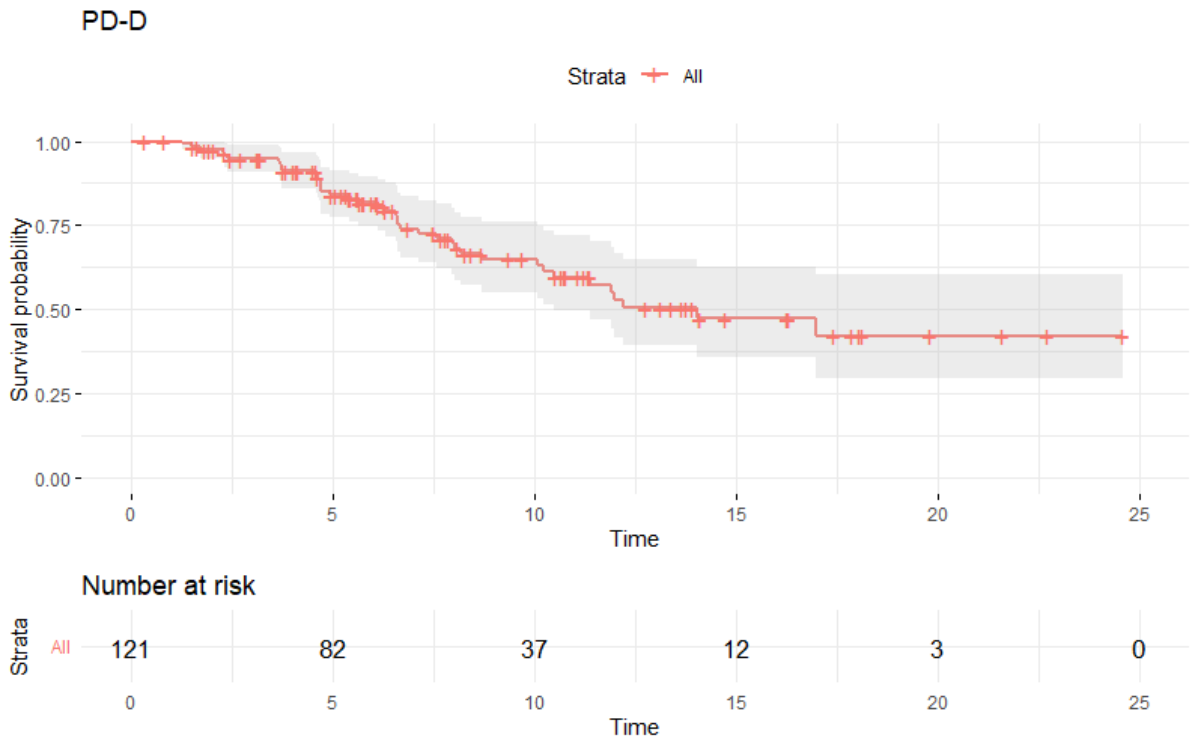

FIGURE 59 – SURVIVAL CURVE OF PARKINSON'S DISEASE PATIENTS OVER TIME THAT DEVELOP MILD COGNITIVE IMPAIRMENT (PDMCI)

Survival plot for all PD patients from the discovery study and their survival probability of declining to a Montreal Cognitive Assessment (MoCA) score of <21 over time. Time scale shows years from PD diagnosis. Trendline shows mean survival probability and 95% confidence interval.

TABLE 21 – COX REGRESSION OF PARKINSON'S DISEASE DEVELOPING DEMENTIA LEVEL COGNITIVE IMPAIRMENT

Cox regression of plasma protein levels in Parkinson's Disease patients versus risk of declining cognitively to Montreal Cognitive Assessment <21, with age and sex as covariates. Columns show protein gene name, Cox coefficient, Hazard ratio, standard error, z-statistic, and p-value.

		Conversion to MoCA <21 (equivalent to PD-D)			
Gene	*coef*	*Hazard ratio exp(coef)*	*se(coef)*	*z*	*p-value*
OBSCN	-3.645	0.026	0.982	-3.711	0.00021
TERF2	-3.201	0.041	0.930	-3.442	0.00058
UFL1	-3.942	0.019	1.151	-3.426	0.00061
GFPT2	-2.406	0.090	0.707	-3.402	0.00067
PROS1	-6.410	0.002	1.924	-3.332	0.00086
PREX2	-6.021	0.002	1.818	-3.312	0.00093
G3BP2	-2.875	0.056	0.876	-3.284	0.00102
RYR3	-3.502	0.030	1.086	-3.225	0.00126
H2AW	-6.594	0.001	2.060	-3.201	0.00137
FASLG	-4.786	0.008	1.501	-3.189	0.00143
MYO18A	2.385	10.862	0.751	3.175	0.00150
KMT2D	3.135	22.987	0.991	3.163	0.00156
CCDC150	-2.838	0.059	0.918	-3.091	0.00199
LAMA3	5.318	203.991	1.725	3.084	0.00204
PDE6B	3.631	37.764	1.203	3.018	0.00255
RNF213	-3.683	0.025	1.224	-3.010	0.00262
COL11A2	-4.447	0.012	1.479	-3.007	0.00263
TAOK2	-3.285	0.037	1.094	-3.003	0.00267
JPH3	-2.576	0.076	0.861	-2.991	0.00278
C5	-6.903	0.001	2.344	-2.945	0.00323
UBR1	-3.424	0.033	1.171	-2.923	0.00347
DSCAM	-3.907	0.020	1.343	-2.909	0.00363
CBX3	8.604	5455.766	2.961	2.906	0.00366
PRR14L	-4.872	0.008	1.685	-2.891	0.00384
MGAT5	-2.194	0.111	0.759	-2.890	0.00386
CASZ1	-5.225	0.005	1.815	-2.879	0.00399
IBTK	2.224	9.248	0.777	2.861	0.00422
SYT13	-3.688	0.025	1.289	-2.860	0.00423
ERICH3	-2.339	0.096	0.821	-2.848	0.00440
CPN2	6.099	445.291	2.152	2.834	0.00460

Cognitive decline in PD is believed to begin with subjective cognitive decline, followed by mild cognitive decline, and finally dementia. PD-MCI is seen as a prodromal stage of PDD, and a major risk factor. Hence, individuals that develop MCI rapidly also tend to decline further and develop dementia. It therefore stands to reason the proteins which reflect conversion to PDMCI, also reflect PDD conversion. Therefore, the proteins that were significant for both models in Table 22 were summarised, and many of the top proteins did indeed overlap.

Optimal cut-off points were established for the top 3 proteins from the cox regressions, Vitamin K-dependent protein S, Telomeric repeat-binding factor 2, and Obscurin, and survival curves were plotted for each of them as a visual representation of how well they predicted development of PDMCI and PDD respectively (Figure 60).

Visually, all three of these proteins appear to discriminate well between PD patients that decline fast and slow cognitively, both when the threshold is set to when they develop dementia and when they develop MCI. Moreover their hazard ratio is quite large (Table 22), which makes them promising biomarker candidates differentiate between fast and slow cognitively declining patients.

If a protein level reflects cognitive decline in an individual with PD, it is likely that it also reflects cognitive ability at a cross sectional level at the same time. Therefore, it was investigated which proteins would reflect both cognitive performance as well as cognitive decline (Table 23).

TABLE 22 – COX REGRESSION OF PARKINSON'S DISEASE DEVELOPING MILD COGNITICE IMPAIRMENT OR DEMENTIA – OVERLAPPING PROTEINS

Cox regression of plasma protein levels in Parkinson's Disease patients versus risk of declining cognitively to Montreal Cognitive Assessment <26 and <21, with age and sex as covariates. Columns show protein gene name, Cox coefficient (coef), Hazard ratio (exp(coef)), standard error (se(coef)), and p-value. Proteins with p-value <0.01 for both conversion to PDMCI and PDD are shown.

		Conversion to MoCA <26 (equivalent to PD-MCI)				*Conversion to MoCA <21 (equivalent to PD-D)*			
Protein	*Gene*	*coef*	*exp(coef)*	*se(coef)*	*p-value*	*coef*	*exp(coef)*	*se(coef)*	*p-value*
Vitamin K-dependent protein S	*PROS1*	-5.52	0.004	1.55	0.00036	-6.41	0.002	1.92	0.00086
Telomeric repeat-binding factor 2	*TERF2*	-2.59	0.075	0.74	0.00043	-3.20	0.041	0.93	0.00058
Obscurin	*OBSCN*	-2.81	0.060	0.81	0.00051	-3.64	0.026	0.98	0.00021
E3 UFM1-protein ligase 1	*UFL1*	-2.50	0.082	0.81	0.00210	-3.94	0.019	1.15	0.00061
Lumican	*LUM*	-3.92	0.020	1.29	0.00230	-5.02	0.007	1.78	0.00470
Collagen alpha-2(XI) chain	*COL11A2*	-3.46	0.031	1.16	0.00295	-4.45	0.012	1.48	0.00263
Acetyl-CoA carboxylase 1	*ACACA*	2.40	11.033	0.84	0.00444	3.30	27.004	1.24	0.00764
E3 ubiquitin-protein ligase RNF213	*RNF213*	-2.73	0.065	0.96	0.00454	-3.68	0.025	1.22	0.00262
Ectonucleoside triphosphate diphosphohydrolase 7	*ENTPD7*	1.99	7.344	0.74	0.00680	2.34	10.394	0.86	0.00670
Down syndrome cell adhesion molecule	*DSCAM*	-2.97	0.051	1.12	0.00796	-3.91	0.020	1.34	0.00363
Serine/threonine-protein kinase TAO2	*TAOK2*	-2.33	0.098	0.88	0.00836	-3.28	0.037	1.09	0.00267
Biotinidase	*BTD*	-3.38	0.034	1.28	0.00843	-4.18	0.015	1.57	0.00794
PH and SEC7 domain-containing protein 3	*PSD3*	2.52	12.407	0.96	0.00850	3.45	31.557	1.32	0.00884
Chromobox protein homolog 3	*CBX3*	5.21	183.841	1.99	0.00876	8.60	5455.766	2.96	0.00366

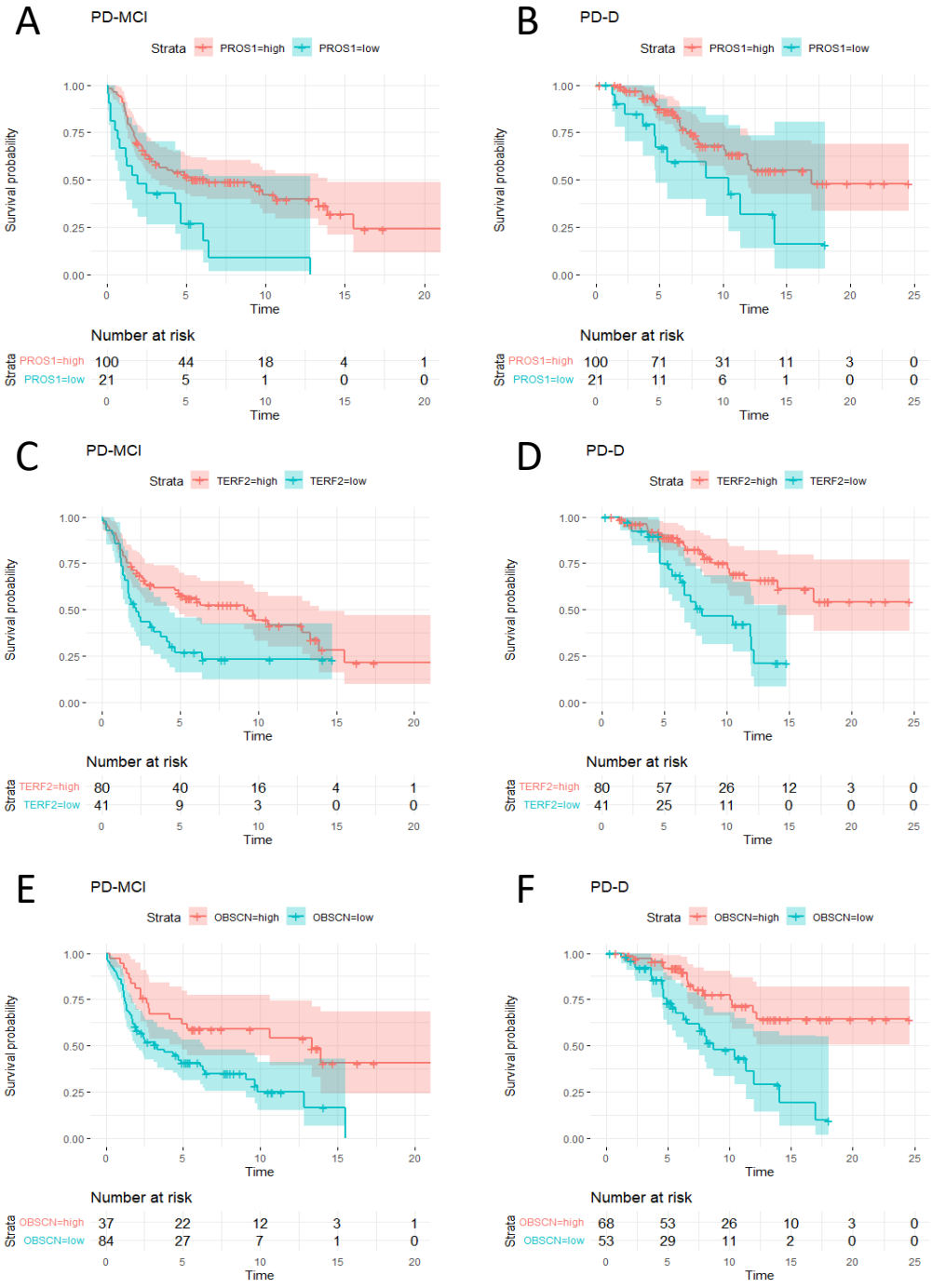

FIGURE 60 – SURVIVAL CURVES OF TOP 3 CANDIDATES THAT PREDICT COGNITIVE DECLINE IN PARKINSON'S DISEASE

Cox regression survival curves for (A,B) Vitamin K-dependent protein S (PROS1), (C,D) Telomeric repeat-binding factor 2 (TERF2), (E,F) and Obscurin (OBSCN), with risk for Parkinson's Disease (PD) patients to develop Mild Cognitive Impairment (MCI) equivalent Montreal Cognitive Assessment (MoCA) <26 or Dementia equivalent MoCA <21 over time. High and Low protein levels set by optimal cut-off. Trendline shows mean survival probability and 95% confidence interval.

TABLE 23 – COMBINED TOP CANDIDATE MARKERS FOR PARKINSON'S DISEASE (PD) COGNITIVE DECLINE AND CROSS SECTIONAL COGNITIVE IMPAIRMENT

Plasma protein levels that both reflect risk of Parkinson's Disease to develop mild cognitive impairment (MCI) and dementia (D) from a cox regression model, as well as cross sectional cognitive performance level in PD patients using Montreal Cognitive Assessment (MoCA). Age and sex were used as covariates, p<0.05 considered significant.

	MoCA correlation		*PD-MCI survival*		*PD-D survival*	
Gene	*t-value*	*p-value*	*coef*	*p-value*	*coef*	*p-value*
CD74	-3.16	0.0023	3.89	0.0084	4.47	0.0186
PROS1	3.05	0.0028	-5.52	0.0004	-6.41	0.0009
BTD	3.03	0.0033	-3.38	0.0084	-4.18	0.0079
JPH3	2.75	0.007	-1.66	0.0178	-2.58	0.0028
SMARCAD1	2.77	0.0072	-2.68	0.0234	-3.71	0.028
C2CD5	-2.73	0.0075	3.97	0.0007	3.57	0.0275
APOA4	2.63	0.0095	-2.40	0.0149	-3.32	0.0063
BOD1L1	-2.6	0.0103	1.60	0.0180	1.83	0.0365
ERICH3	2.5	0.0142	-1.55	0.0301	-2.34	0.0044
PRR14L	2.24	0.0279	-2.82	0.0363	-4.87	0.0038
XRCC6	-2.21	0.0300	3.31	0.0066	3.97	0.0113
UBR4	2.17	0.0318	-1.92	0.0259	-3.04	0.0059
TERF2	2.16	0.0324	-2.59	0.0004	-3.20	0.0006
CCDC136	2.14	0.0354	-1.96	0.0153	-2.29	0.0101
KMT2D	-2.12	0.0364	1.96	0.0179	3.14	0.0016
KMT2B	2.12	0.0365	-1.57	0.0407	-2.91	0.0092
ARMCX1	2.05	0.0441	-3.29	0.0066	-3.03	0.0231
C5	2.03	0.0448	-4.28	0.0119	-6.90	0.0032
UFL1	2.02	0.0455	-2.50	0.0021	-3.94	0.0006
MYO15A	-1.99	0.0492	2.59	0.0271	3.30	0.0242

It was noticed that many of the proteins that correlated with MoCA score, also reflected cognitive decline over time. As these overlapping proteins reflect cognitive impairment both at the time of blood draw as well as over time, they are likely to reflect underlying pathology associated with cognitive function in PD patients.

The sex-specific analysis was also performed for the cox regressions for cognitive decline. Again, a larger number of proteins significantly correlating with cognitive score was seen in the male PD cohort. This could be attributed to the nature of the male PD cohort, which has a larger number of patients, and a greater distribution of MoCA scores. Many of the proteins that were associated with cognitive

decline in the whole PD cohort, such as LUM, PROS1, and TERF2, were predominantly associated with cognitive decline in male PD patients. The results are summarised in Table 24.

Individual peptides were analysed from the proteins that were significantly associated with conversion to both PDMCI and PDD. The ceruloplasmin (CP) peptide IYHSHIDAPK and the hemopexin (HPX) peptide GEVPPR were associated with lower risk of developing cognitive impairment and dementia ($p < 0.05$), and the Glycogen Phosphorylase L (PYGL) peptide ISLSNESNK was associated with higher risk of developing cognitive impairment and dementia ($p < 0.05$).

TABLE 24 – COX REGRESSION OF PARKINSON'S DISEASE DEVELOPING MILD COGNITICE IMPAIRMENT OR DEMENTIA – SEX-SPECIFIC ANALYSIS

Cox regression of plasma protein levels in Parkinson's Disease patients, females and males separately, versus risk of declining cognitively to Montreal Cognitive Assessment <26 and <21, with age and sex as covariates. Columns show protein gene name, Hazard ratio (exp(coef)), and p-value. Top results for females and males separately, for both conversion to PDMCI and PDD are shown.

	Females PDMCI conv		*Females PDD conv*			*Males PDMCI conv*		*Males PDD conv*	
Gene	*HR*	*p-value*	*HR*	*p-value*	*Gene*	*HR*	*p-value*	*HR*	*p-value*
PCNX1	3160.121	0.00014	1122.055	0.05412	*MYO18A*	200.2177	0.00001	590.6372	0.00002
PHF21A	0.000305	0.00015	0.002178	0.26949	*LUM*	0.000101	0.00001	2.45E-05	0.00004
ROCK2	0.005107	0.00059	0.002197	0.00273	*PROS1*	0.000417	0.00007	0.00052	0.00100
B4GALNT4	62.22995	0.00091	53.95796	0.08873	*TERF2*	0.026322	0.00008	0.013219	0.00007
KDM3B	279.4508	0.00102	4104.23	0.10471	*DSCAM*	0.000572	0.00011	0.000368	0.00054
PRRC2C	162.0425	0.00118	386.6055	0.03879	*ANK2*	36.64035	0.00026	192.1883	0.00008
ARHGAP23	1337.425	0.00124	203.1372	0.08725	*UFL1*	0.015625	0.00030	0.001025	0.00017
ZNF407	0.02297	0.00147	0.048524	0.05700	*COL11A2*	0.005711	0.00036	0.005282	0.00241
ODF2	0.002177	0.00240	0.436468	0.86242	*OBSCN*	0.018573	0.00039	0.00842	0.00065
IQCC	0.018659	0.00391	0.000583	0.03211	*SERPING1*	0.000721	0.00046	0.000274	0.00163
RIC1	1416.898	0.00437	19542.6	0.01312	*NEK3*	0.016137	0.00053	0.007317	0.00077
IGKV2-28	15.17409	0.00449	9.172528	0.08190	*EMILIN2*	173864	0.00059	7260141	0.00137
STYXL2	0.012078	0.00492	0.325606	0.69442	*INVS*	7.99E-05	0.00069	0.000422	0.01484
UBN2	4645.712	0.00504	40.90032	0.25928	*PREX2*	0.001984	0.00070	0.000269	0.00068
V162	21.62524	0.00578	8.78359	0.17520	*LRG1*	0.004893	0.00071	0.013069	0.03259
ACSBG2	394.0366	0.00638	27.79217	0.29089	*DZIP1*	25.84563	0.00073	8.293835	0.08252
SIPA1L1	60.06817	0.00645	0.993944	0.99793	*PTPRN*	0.000923	0.00073	0.000146	0.00242
AGAP2	0.005035	0.00672	0.007086	0.08431	*IDS*	0.011989	0.00077	0.003677	0.00070
CCNT2	0.041591	0.00699	0.037453	0.22852	*ADGRB3*	485.46	0.00090	1340.573	0.00110
SBNO1	27.87462	0.00718	2.672091	0.53533	*AHNAK2*	60.26208	0.00093	90.16858	0.00364
FAT4	0.044613	0.00773	0.136077	0.22207	*SYT13*	0.00953	0.00104	0.005761	0.00365
GAD1	153.2922	0.00787	43.93959	0.15609	*PROC*	0.005779	0.00106	0.027302	0.07585
PRKG2	0.030978	0.00834	0.101807	0.38207	*TCF20*	0.036194	0.00128	0.044048	0.01302
CUX2	0.032675	0.00873	0.058919	0.07063	*BTD*	0.000809	0.00178	1.11E-05	0.00148
DSCAML1	0.079254	0.00882	0.065803	0.08045	*ARMCX1*	0.007314	0.00184	0.005366	0.00491
COL5A2	1192.963	0.00908	0.051329	0.54800	*CACNA1D*	298.502	0.00189	36.01278	0.11141
EFCAB5	876.8873	0.01045	165564.1	0.00690	*ACACA*	25.00983	0.00196	89.64027	0.00364
XRCC6	227.6774	0.01054	9.287929	0.27912	*MMP1*	0.069282	0.00198	0.033476	0.00411
IQGAP3	0.04874	0.01097	0.093837	0.17807	*SNRPD2*	0.02461	0.00199	0.015585	0.01373
V2242	0.0344	0.01132	0.108416	0.29899	*HPX*	0.000453	0.00208	0.000277	0.00700

To summarise, several plasma proteins were associated with cognition in PD patients. The proteomic data was analysed with cognitive decline both cross sectionally and longitudinally. The PD patients were subdivided into those without dementia, those with mild cognitive impairment, and those with dementia, based on conventional cut-off. There was no apparent separation in the proteome data using PCA, and t-SNE plots, and individual plasma proteins that were be associated with cognition were analysed. Following cross-sectional analysis using age, sex, and disease duration as covariates, the top hits were ranked by p-value, and highlighted the ones that were significantly influenced by demographic covariates.

Next, proteins associated with earlier conversion to cognitive impairment and dementia after PD diagnosis were studied. Many proteins were significantly associated with both faster decline to cognitive impairment and dementia after correcting for age and sex. Interestingly, these candidate markers also displayed high hazard ratio. These proteins could therefore potentially be good predictors for cognitive decline if validated in a newly diagnosed PD cohort. Moreover, 20 candidate markers were significant for both cross sectional cognitive score, as well as longitudinal cognitive decline.

Sex specific analysis revealed several differences in proteins associated with PD cognitive impairment and decline, with a larger number of significant results for the male patients.

The top candidate biomarkers for both cognitive performance and cognitive decline were selected for the verification study.

3.2 Verification study

The discovery study yielded many candidate markers that either reflected PD diagnosis, PD motor severity, cognitive impairment, or reflected cognitive decline. The protein IDs used in the analysis were found at a 5% FDR proteome database matching, which warrants verification and validation experiments to gain more confidence in the results. Moreover, the samples were heavily manipulated prior to analysis (depletion, fractionation, and labelling), which introduced many technical covariates. Although many of these variables were controlled for, the final peptide sample analysed on the mass spectrometer was several processing steps away from whole blood plasma. However, proceeding straight to an antibody-based validation method, such as an immunoassay, would introduce a large technical discrepancy, where new technical variables such as antibody binding, antibody specificity, matrix effects etc. would be introduced. Rather than a validation study, a technical verification step on the mass spectrometer is used here, where only high confidence peptides are quantified with a targeted approach. Once the peptides levels are verified and are deemed reliably differentially expressed in PD plasma samples, one can proceed with validating the candidate markers with an immunoassay. Hence, Parallel Reaction Monitoring (PRM) was used, a targeted mass spectrometric technique where several specific peptides from several proteins can be simultaneously monitored and quantified in samples. Moreover, to avoid technical covariates as wherever possible, it was decided to run peptide sample as neat as possible, without any protein depletion, labelling, or fractionation of the samples as was done for the discovery study.

The verification study overview is shown in Figure 61. In brief, a mixed plasma sample was used to set up the method and identify which peptides were good candidates and could be repeatedly quantified. Once these peptides were selected, the study samples, along with Quality Control samples every 10 injections were analysed on the mass spectrometer. Similar to in the discovery study, several pre-processing steps were implemented on the raw data before ending up with the final dataset that was used for analysis. Finally, it was investigated how well the verified proteins correlated with the results from the discovery study as well as their performance as biomarkers.

3.2.1 Study overview

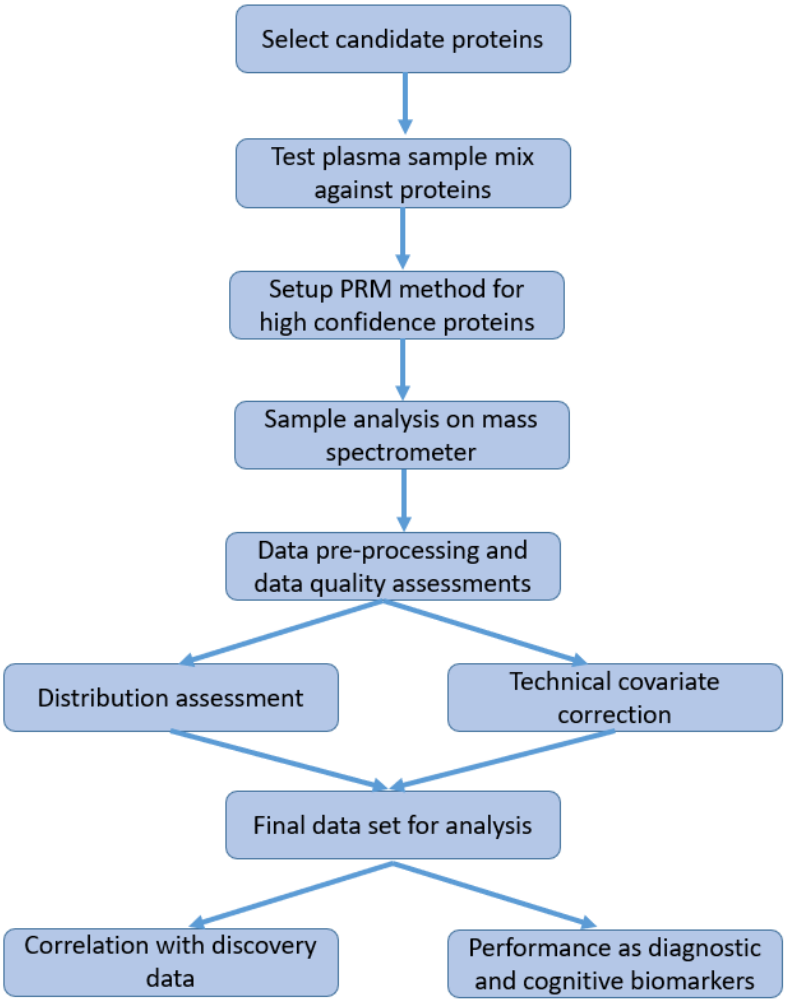

FIGURE 61 – WORKFLOW SCHEMATIC OF VERIFICATION STUDY

3.2.2 Method optimisation

Once the candidate biomarkers were identified in the discovery study, they needed to be narrowed down to a smaller set to be validated on the mass spectrometer using PRM. The top 70 candidates from the discovery study were selected, half of which were from the PD diagnosis and motor severity analyses, and half were selected from the cognition performance and cognitive decline analyses. A plasma mix sample composed of 8 PD samples and 8 healthy controls were used for the PRM setup and testing. One neat sample and one sample depleted from Albumin and IgG were used for the testing.

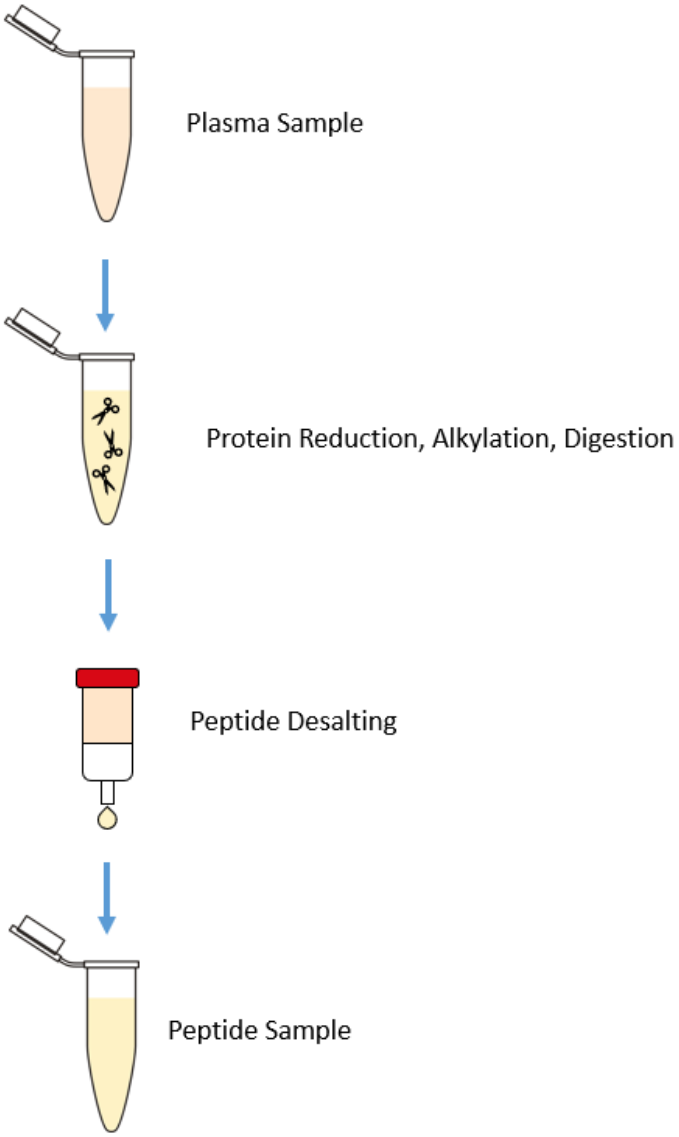

FIGURE 62 – SCHEMATIC OVERVIEW OF SAMPLE PREPARATION FOR THE VERIFICATION STUDY

The method optimisation on the mass spectrometer was performed by the Centre of Excellence for Mass Spectrometry (CEMS)-Denmark Hill Proteomics Facility. Peptides were selected if they were unique to the protein, within the mass range of the instrument, fully tryptic, not susceptible to amino acid modifications, and had doubly or triply charged precursor ions [138]. 12 of the candidate proteins had at least 1 high confidence peptide that was repeatedly and reliably quantified. These were Apolipoprotein A-IV, Apolipoprotein C-III, Complement component C9, Plastin-2, Lumican, Xaa-Pro dipeptidase, Vitamin K-dependent protein S, Serum amyloid A-1 protein, L-selectin, Transforming growth factor-beta-induced protein ig-h3, Thrombospondin-1, and E3 ubiquitin-protein ligase TRIM33. The details on these proteins from the discovery study are summarised in Table 25. The list of peptides is shown in Table 26. It was noticed that most detected proteins, except for TRIM33, had a high confidence in the database matching in the discovery study. There was no difference in which peptides could be quantified with PRM between the albumin and IgG depleted sample and the undepleted sample. Therefore, it was decided not to deplete samples from albumin and IgG, for the quantified protein levels in the sample best reflect the relative protein composition of whole plasma.

TABLE 25 – THE 12 QUANTIFIABLE CANDIDATE BIOMARKER PROTEINS USING PARALLEL REACTION MONITORING ON PLASMA, PROTEIN DATA FROM DISCOVERY STUDY

Table showing the 12 biomarker candidates from the discovery study that could reliably be quantified using parallel reaction monitoring. Columns show protein information from the discovery study, including Gene name, False Discovery Rate (FDR) confidence estimation, Accession number, q-value, Protein coverage %, number of peptides identified, number of peptide spectral matches (PSMs), number of protein unique peptides, and results from the biomarker analysis.

Protein name	*Gene*	*FDR Conf.*	*Access-ion*	*max q-value*	*Coverage %*	*n Peptides*	*n PSMs*	*n Protein Unique Peptides*	*Biomarker purpose*
Apolipoprotein A-IV	*APOA4*	High	P06727	0	100	82	4297	81	Cognitive Decline and Cognitive Impairment
Apolipoprotein C-III	*APOC3*	High	P02656	0	81	18	641	18	PD diagnostic
Complement component C9	*C9*	High	P02748	0	98	86	1828	83	PD diagnostic
Plastin-2	*LCP1*	High	P13796	0.007	79	67	314	62	PD diagnostic
Lumican	*LUM*	High	P51884	0	74	28	571	28	Cognitive Decline
Xaa-Pro dipeptidase	*PEPD*	High	P12955	0.005	72	40	317	40	PD diagnostic
Vitamin K-dependent protein S	*PROS1*	High	P07225	0	81	73	1125	71	Cognitive Decline and Cognitive Impairment
Serum amyloid A-1 protein	*SAA1*	High	P0DJI8	0	82	17	338	8	Motor and Cognitive Impairment
L-selectin	*SELL*	High	P14151	0.005	55	29	232	29	PD diagnostic
Transforming growth factor-beta-induced protein ig-h3	*TGFBI*	High	Q15582	0.005	81	70	319	70	PD diagnostic and Motor Impairment
Thrombospondin-1	*THBS1*	High	P07996	0	87	139	438	135	PD diagnostic
E3 ubiquitin-protein ligase TRIM33	*TRIM33*	Low	Q9UPN9	0.299	61	81	284	81	Motor Impairment

TABLE 26 – THE 12 QUANTIFIABLE CANDIDATE BIOMARKER PROTEINS USING PARALLEL REACTION MONITORING ON PLASMA, LIST OF PEPTIDES

Table showing the 12 biomarker candidates from the discovery study that could reliably be quantified using targeted mass spectrometry, with the peptides for each protein that could be reliably quantified.

Protein	*Gene name*	*Peptide*
Apolipoprotein A-IV	*APOA4*	EAVEHLQK
		IDQNVEELK
		IDQTVEELR
		ISASAEELR
		LEPYADQLR
Apolipoprotein C-III	*APOC3*	DALSSVQESQVAQQAR
Complement component C9	*C9*	AIEDYINEFSVR
		DVVLTTTFVDDIK
		TEHYEEQIEAFK
Plastin-2	*LCP1*	VYALPEDLVEVNPK
Lumican	*LUM*	ILGPLSYSK
		SLEDLQLTHNK
Xaa-Pro dipeptidase	*PEPD*	AVYEAVLR
Vitamin K-dependent protein S	*PROS1*	NNLELSTPLK
		SQDILLSVENTVIYR
Serum amyloid A-1 protein	*SAA1*	FFGHGAEDSLADQAANEWGR
		GPGGVWAAEAISDAR
L-selectin	*SELL*	AEIEYLEK
Transforming growth factor-beta-induced protein ig-h3	*TGFBI*	QHGPNVCAVQK
		VLTDELK
Thrombospondin-1	*THBS1*	GTLLALER
E3 ubiquitin-protein ligase TRIM33	*TRIM33*	GAIENLLAK

3.2.3 PRM data pre-processing

Once the peptides were selected, all the study samples were processed and run on the mass spectrometer. A QC sample along with a blank sample was injected after each 10 study samples. All samples were run in duplicate. Peptide and fragment ion peaks were identified in the Skyline software automatically (and double checked manually), and peak areas were exported.

The pre-processing stages of the raw output data were similar to the discovery study. Firstly, the protein distributions (mean peptide peak area) of each of the 12 targeted proteins were visualised (Figure 63). A right skew of the quantification data was observed for all proteins, which looked more normally distributed after log10 transforming the data. This is in line with what was observed in the discovery data.

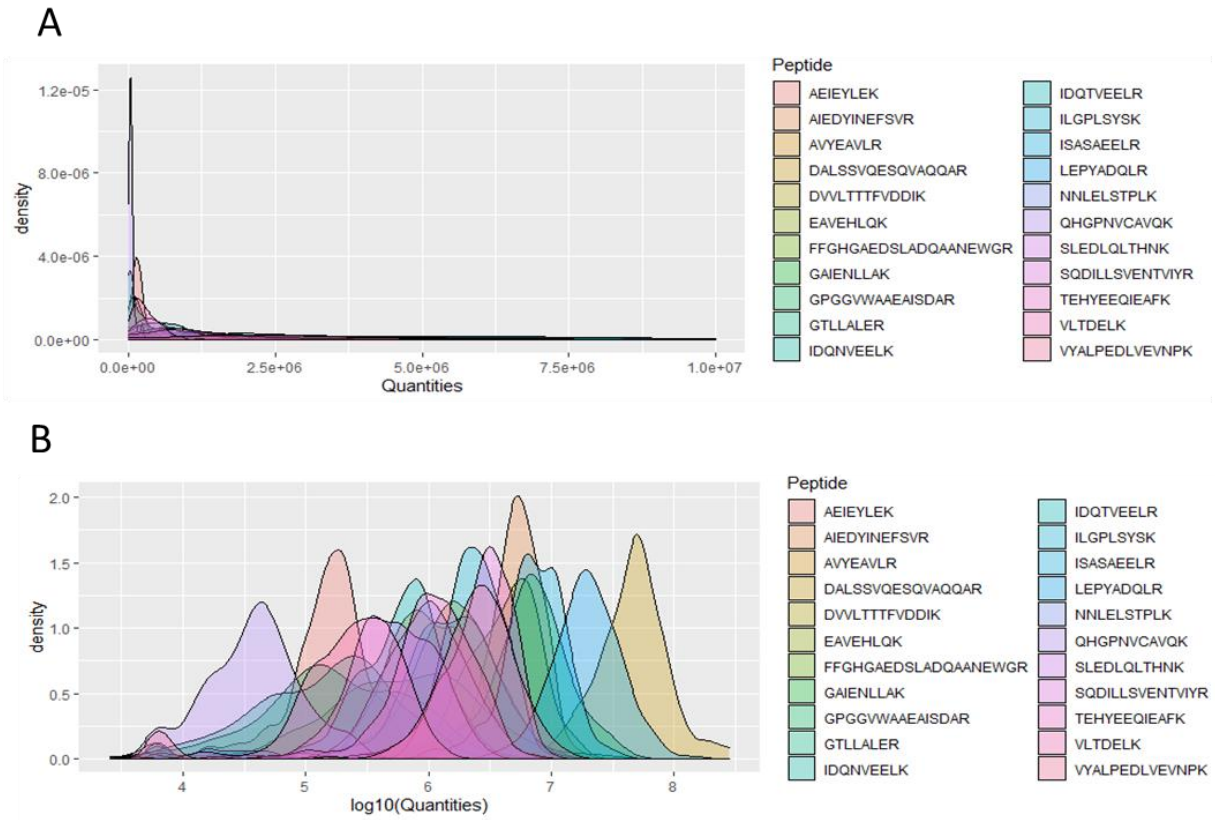

FIGURE 63 – PROTEIN LEVEL DISTRIBUTIONS OF THE 12 QUANTIFIED PROTEINS USING PARALLEL REACTION MONITORING

A) Density plots representing distribution of protein levels across all samples for each of the 12 proteins quantified using parallel reaction monitoring (PRM). B) Log-transformed protein levels across all samples for each of the 12 proteins quantified using parallel reaction monitoring (PRM).

When performing the data quality assessment, a change in signal intensity was observed over time when the samples were run on the mass spectrometer. Figure 64A shows the variation in quantified values against the run order over time, and the change was observed in all samples, including quality controls. This was closer examined in the Skyline software (Figure 64B), where 5 technical batches were identified based on minor shifts in retention time and peak areas. This shift generally did not appear to alter which fragment ion peaks that were quantified, nor their relative abundance to each other. It was however noticed that the peak area decreased with run order both for replicate 1 (first half of samples) and replicate 2 (second half of samples). Around 80 of the samples for the second replicate, labelled as Batch 3/Replicate 2b were analysed again for replicate 2 owing to their exceeding low peak areas compared with the rest of the samples.

To further investigate this shift over time for the two replicates, the log transformed peak areas were plotted against run order, both for all samples and quality controls separately, with the observed technical batches were highlighted (figure 65). It was noticed that there indeed was a general drop in quantified peak area over time for both replicate 1 (batch 1 and 2) and replicate 2 (batch 4 and 5), with a marked difference with lower quantified peak areas for batch 3. This was further confirmed when only plotting QC samples versus run order, as these samples should theoretically have identical peptide levels (figure 65B).

It was investigated whether the change in peptide abundances was due to total peptide abundance, as this was a technical variable responsible for variation in peptide quantities in the discovery study. The total ion count was plotted versus run order, which showed a similar trend to the sample quantifications (figure 66). Hence, all peak areas were corrected to the total ion count for each sample. This would additionally result in a similar correction to what was performed in the discovery study, where the total protein abundance across all samples was normalised and make the two data sets more comparable.

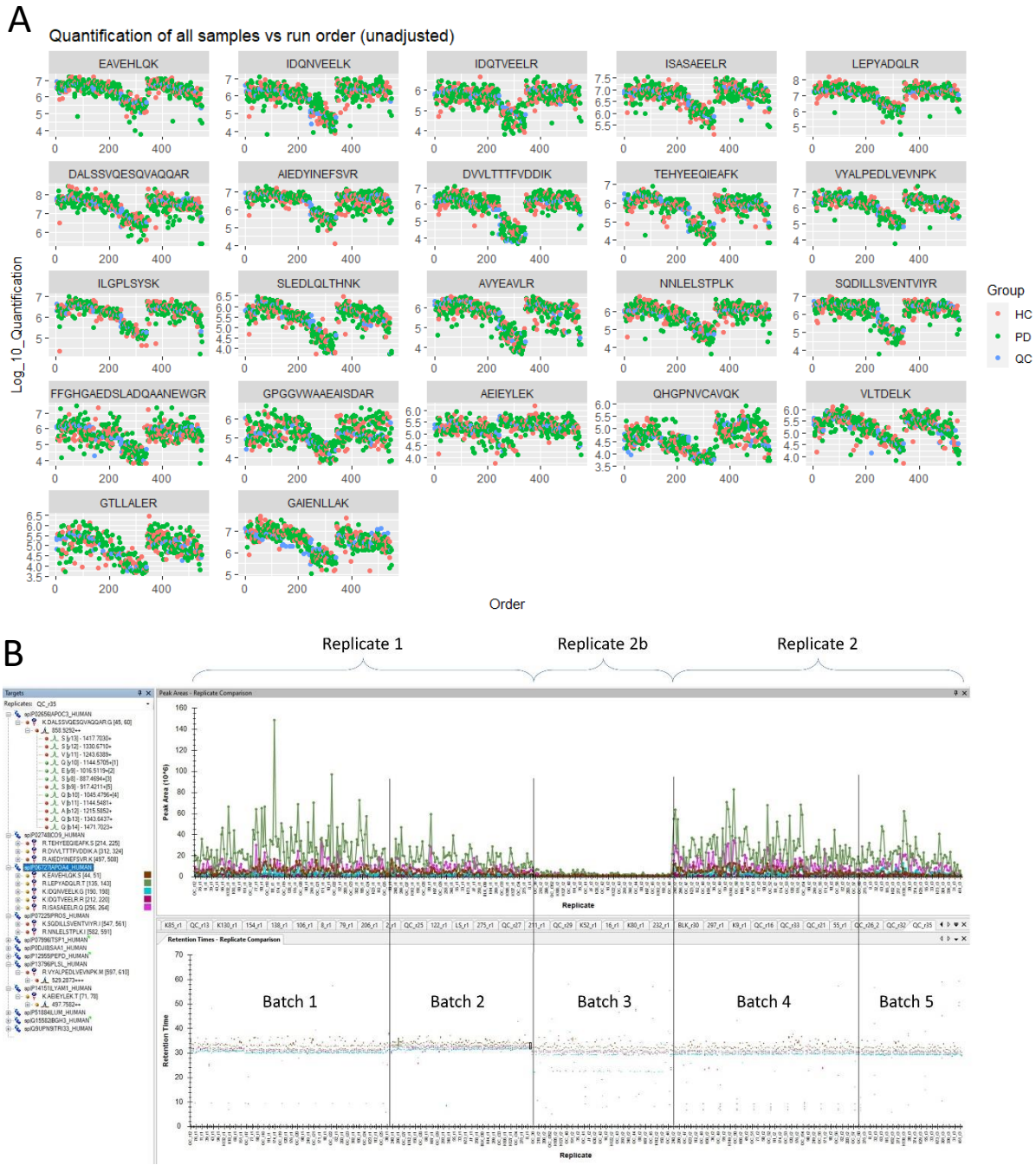

FIGURE 64 – TECHNICAL BATCH EFFECTS AND DIMINISHING PEAK AREAS OVER TIME IN PARALLEL REACTION MONITORING

A) Scatterplots of log transformed peptide abundances versus run order. All samples, Parkinson's Disease (PD), Healthy Controls (HC), and quality controls (QC) plotted. B) Batch effects and replicates in Skyline software. A representative protein (Apolipoprotein A-IV) displayed. Lower graph showing retention time versus run order, upper graph shows peak areas vs run order. Replicates 1 and 2 indicated. Replicate 2b were poor quality samples that were rerun.

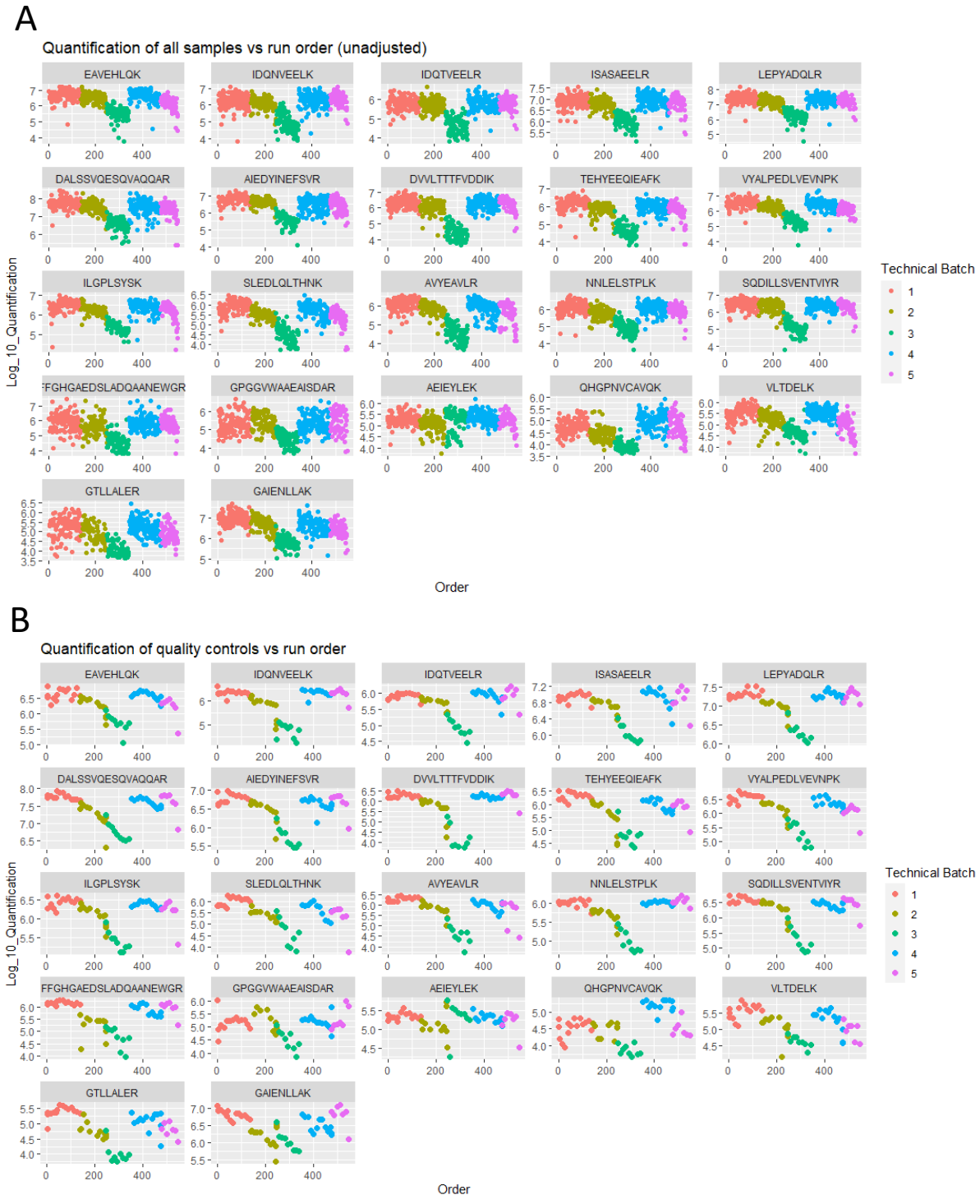

FIGURE 65 – LOG10 TRANSFORMED PEPTIDE PEAK AREAS AGAINST RUN ORDER IN PARALLEL REACTION MONITORING STUDY, TECHNICAL BATCHES INDICATED

A) Scatterplots of log transformed peptide abundances versus run order. All samples, Parkinson's Disease (PD), Healthy Controls (HC), and quality controls (QC), plotted. Observed technical batches highlighted. B) Scatterplots of log transformed peptide abundances versus run order. Quality controls (QC) only plotted. Observed technical batches highlighted.

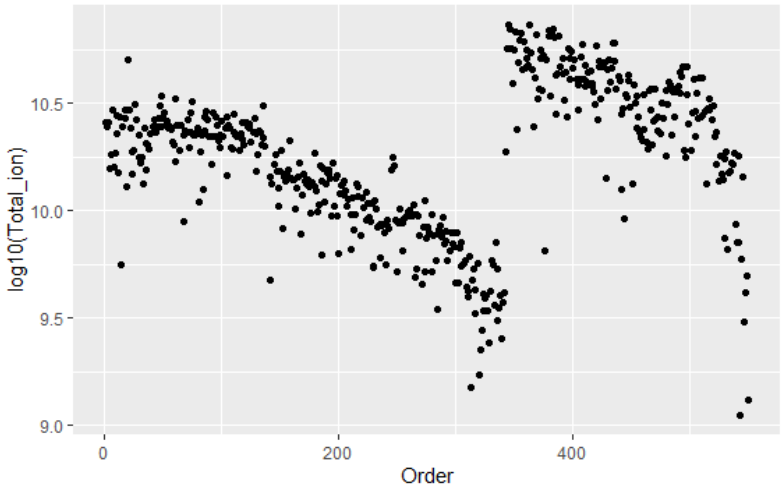

FIGURE 66 – TOTAL ION COUNT AGAINST RUN ORDER

Log10 total ion count for each sample versus run order in the parallel reaction monitoring experiment.

Once all the peak areas were normalised to total ion counts, they were plotted again, and a reduced variation was observed against sample run order, and the dynamic range of each peptide was generally decreased (Figure 67A). Although reduced, there was still a variation in sample abundance against run order and technical batches that needed to be corrected for further. It was noticed the quality control (QC) sample, which was the same sample run every time, displayed variation over time similar to what was observed in all study samples. Therefore, the total ion normalised peak abundances were additionally corrected to the QCs. Unlike the discovery study, an internal QC was not available that was run simultaneously with the study samples. Instead, a QC sample as well as a blank sample was run approximately every 10 samples. A relatively tight-fitting local polynomial regression fitting (loess) curve was fit for all QC samples over time (Figure 67B). The tightness of the loess curve was selected so that the trendline would follow the datapoints closely, without overfitting the curve to outliers. This curve was used to calculate a theoretical QC value for each PD and HC sample as a function of the run order, and all samples were corrected to this estimated intra-QC value. Once the peptides were QC adjusted, they were plotted again with sample run order, which displayed a more even distribution in peptide quantities over time. Both the mean peptide level and sample distribution were more consistent for most of the peptides (Figure 68). A few of the peptides still displayed smaller degree of variation in peak abundance with run order, particularly a variation in the dynamic range was noticed. However, this variation was not consistent across all peptides, and not corrected for further.

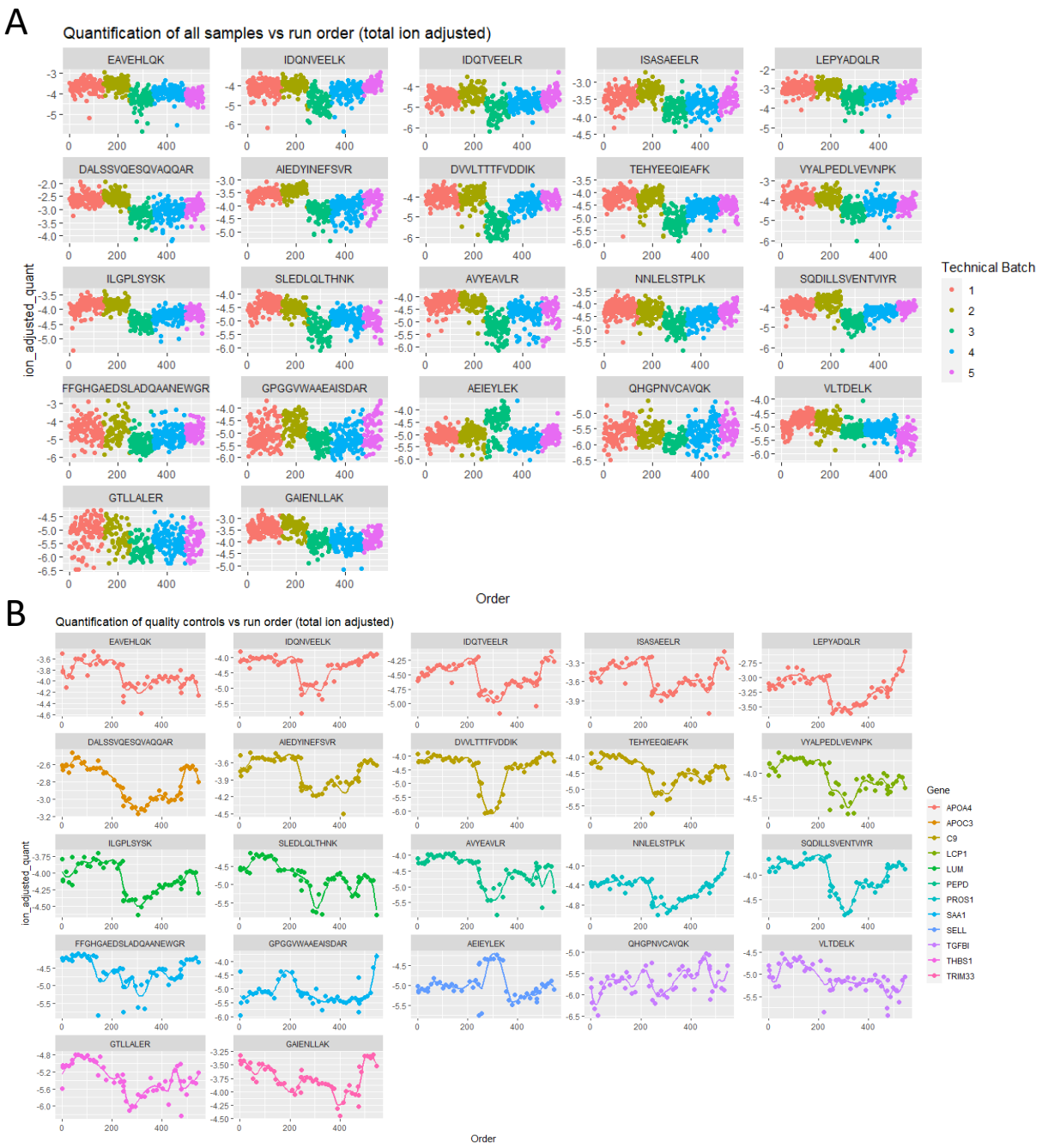

FIGURE 67 – LOG10 TRANSFORMED AND TOTAL ION COUNT ADJUSTED PEPTIDE PEAK AREAS AGAINST RUN ORDER IN THE PARALLEL REACTION MONITORING STUDY

A) Scatterplots of log transformed, and total ion count adjusted peptide abundances versus run order. All samples, Parkinson's Disease (PD), Healthy Controls (HC), and quality controls (QC), plotted. Observed technical batches highlighted. B) Scatterplots of log transformed, and total ion count adjusted peptide abundances versus run order. Quality controls (QC) only plotted. Protein groups indicated. Trendline represents local polynomial regression (loess) curve.

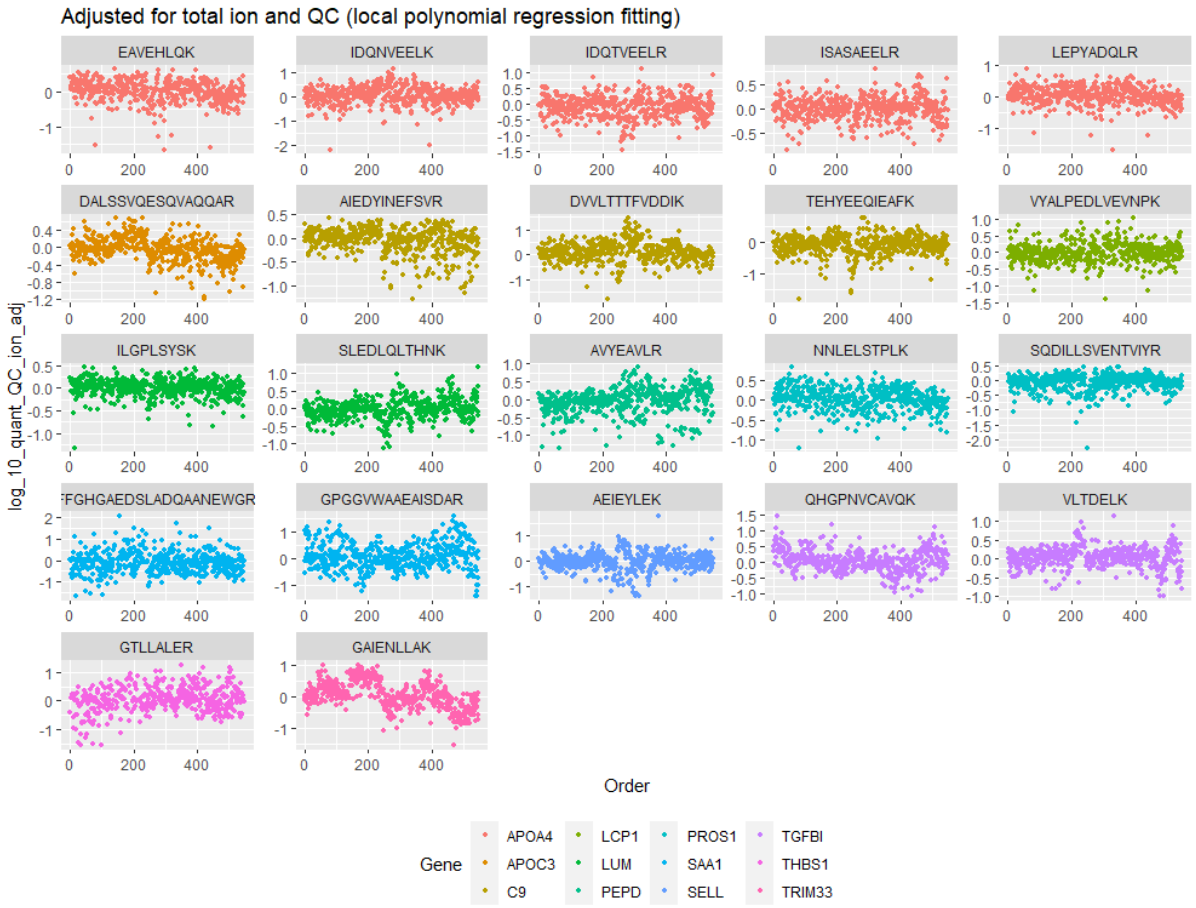

FIGURE 68 – LOG10 TRANSFORMED, TOTAL ION COUNT ADJUSTED, AND QUALITY CONTROL ADJUSTED PEPTIDE PEAK AREAS AGAINST RUN ORDER IN PARALLEL REACTION MONITORING STUDY

Scatterplots of log transformed, total ion count adjusted, and quality control adjusted peptide abundances versus run order. All samples, Parkinson's Disease (PD), Healthy Controls (HC), and quality controls (QC), plotted. Protein groups indicated.

To summarise the PRM raw data pre-processing, the data was log transformed, corrected for total ion count, and adjusted for QC before being used for analysis. One of the technical batches, batch 3, was ultimately eliminated from the dataset. The average from replicate 1 (technical batch 1 and 2), and replicate 2 (batch 4 and 5) was used as the peptide quantity. The two replicates are plotted against each other in Figure 69, to get an overview of how well the peptides in replicates 1 and 2 correlated with each other. It was observed that most peptides correlated well between the two replicates (Spearman $\rho > 0.5$).

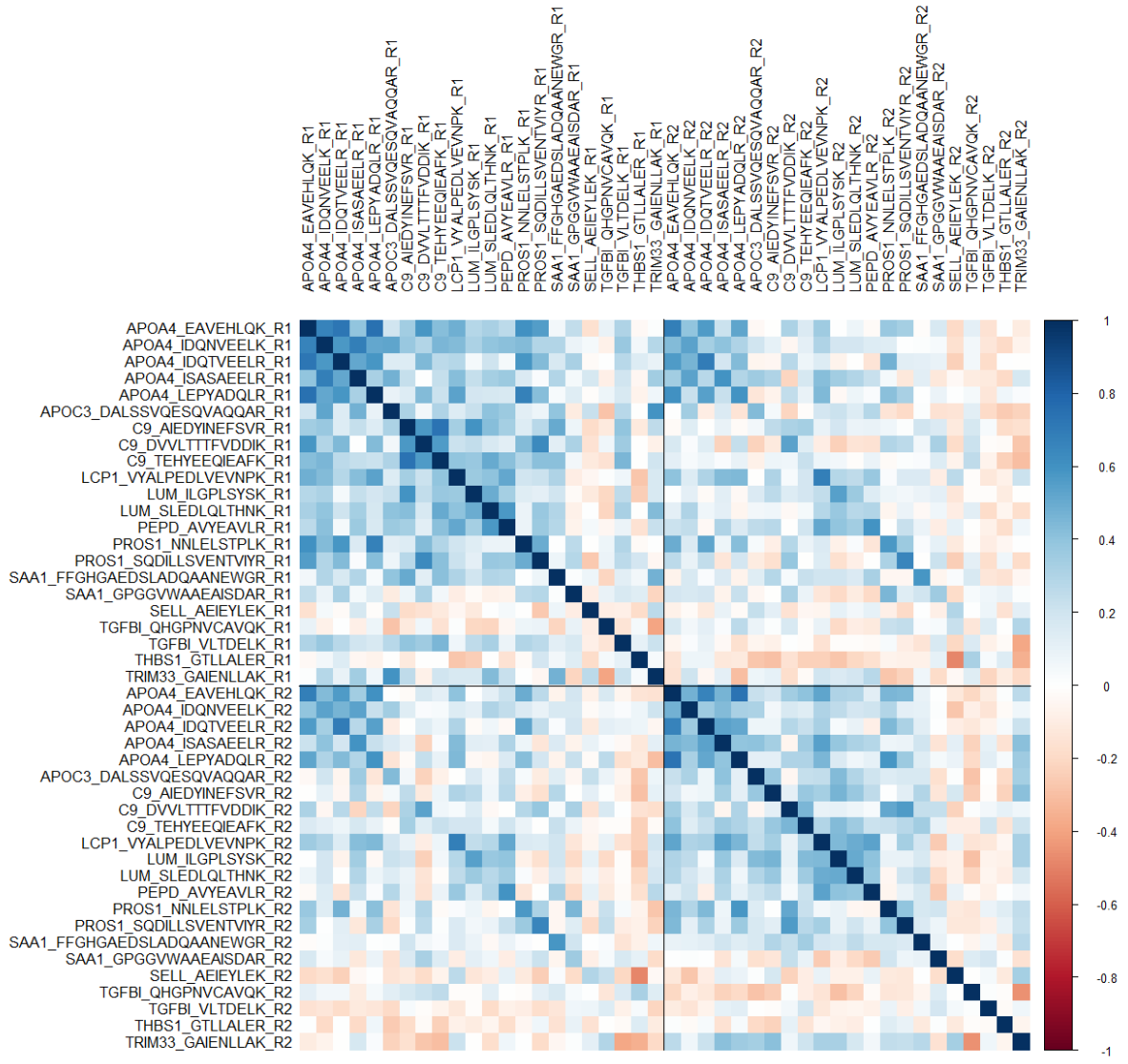

FIGURE 69 – CORRELATION PLOTS BETWEEN PEPTIDE LEVELS BETWEEN THE REPLICATES IN THE VERIFICATION STUDY

Correlation plot with peptide levels from the two replicates from the verification study. Labelled as "gene-name_peptide_replicate". Pearson correlation (r) as colour scale.

3.2.4 PRM data biomarker performance

22 peptides from 12 proteins were successfully detected and quantified using PRM. The peptides were analysed separately as well as averaged for each protein, and analysed against the same clinical parameters as in the discovery study. All peptides were corrected for age and sex in the analyses.

APOC3, C9, LCP1, PEPD, SELL, TGFBI, and THBS1 were all able to differentiate PD from healthy controls with the discovery data. In the PRM data the APOC3 protein (t = -1.99, p = 0.049), the TEHYEEQIEAFK peptide from C9 (t = 0.038, p = 0.017), and the TGFBI protein (t = 2.65, p = 0.0088) were significantly changed in PD compared with HC (Figure 70).

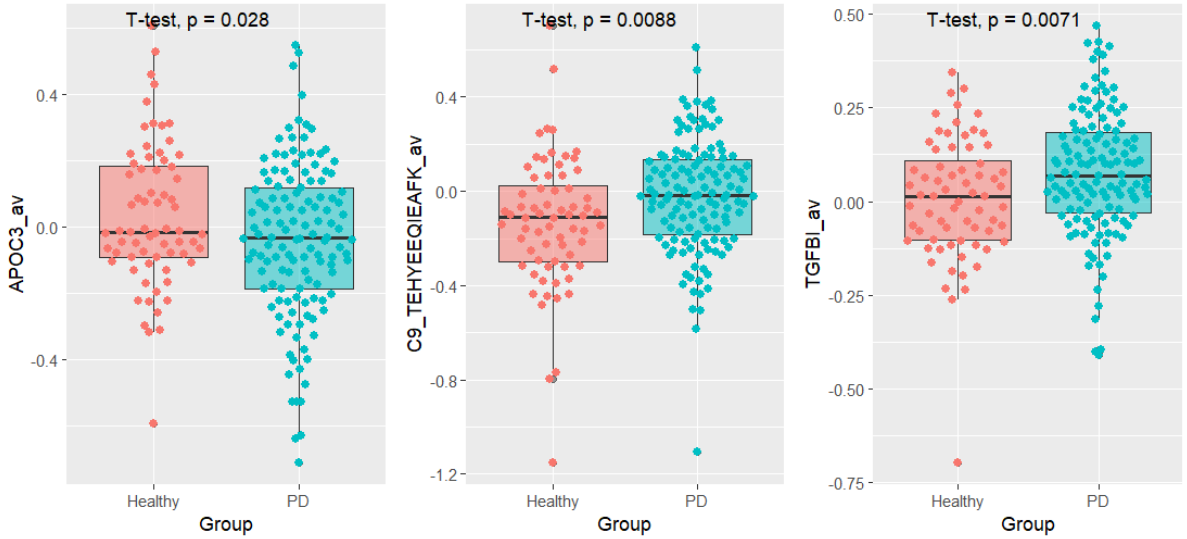

FIGURE 70 BOXPLOTS FOR PLASMA LEVELS OF APOLIPOPROTEIN C3, COMPLEMENT PROTEIN C9 PEPTIDE, AND TRANSFORMING GROWTH FACTOR BETA INDUCED TGFBI BETWEEN PARKINSON'S DISEASE PATIENTS AND HEALTHY CONTROLS

Plasma levels of APOC3, TEHYEEQIEAFK peptide of C9, TGFBI between Parkinson's Disease patients and Healthy Controls. Data from parallel reaction monitoring experiment. Uncorrected t-test with p-value displayed.

SAA1, TGFBI, and TRIM33 all correlated with motor disease severity in the discovery study. Only the FFGHGAEDSLADQAANEWGR peptide of SAA1 was significantly correlated with Hoehn & Yahr (p = 0.016) after correcting for age and sex. APOA4 and SAA1 correlated with cognitive score (MoCA) in the discovery study. Again, only the FFGHGAEDSLADQAANEWGR peptide of SAA1 was significantly correlated with Hoehn & Yahr (p = 0.025) after correcting for age and sex (Figure 71).

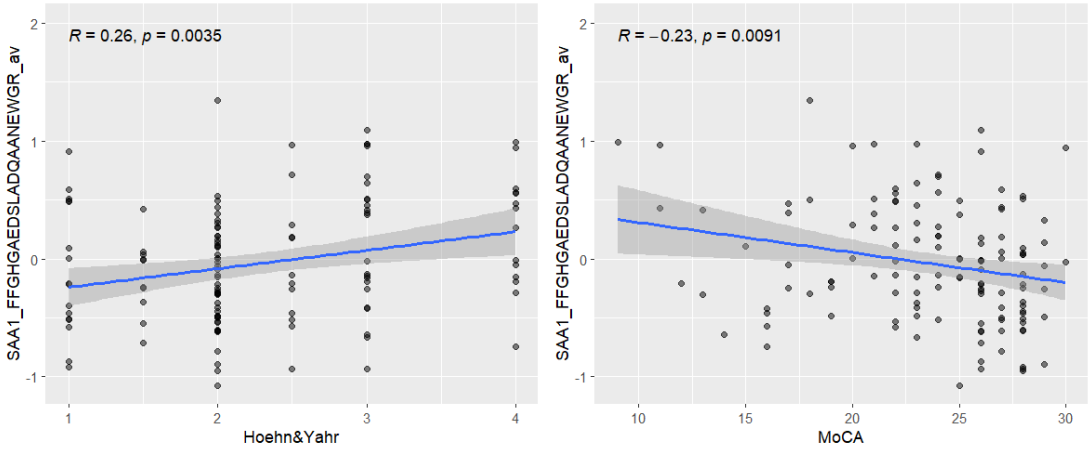

FIGURE 71 – SCATTERPLOT OF SERUM AMYLOID A1 PROTEIN PLASMA LEVELS VERSUS HOEHN & YAHR AND MOCA IN PARKINSON'S DISEASE

Scatterplots of plasma levels of Serum Amyloid A1 (SAA1) FFGHGAEDSLADQAANEWGR peptide in Parkinson's Disease patients versus Hoehn & Yahr and MoCA. Linear trendline with standard error shown. Pearson correlation coefficient and uncorrected p-value shown.

LUM, APOA4, and PROS1 were all associated with cognitive decline and early conversion to PDMCI/PDD in the discovery study. When adding age and sex to the cox regression as covariates, the EAVEHLQK peptide of APOA4 (HR = 0.34, p = 0.016) as well as the protein level of APOA4 (HR = 0.31, p = 0.041) were associated with the conversion to PDD, but not to PDMCI. Both peptides of LUM, ILGPLSYSK (HR = 0.091, p = 0.028) and SLEDLQLTHNK (HR = 0.17, p = 0.028), as well as LUM protein level (HR = 0.085, p = 0.016) were associated with the conversion to PDD but not PDMCI (Figure 72). PROS1 was not associated with cognitive decline in the PRM cohort.

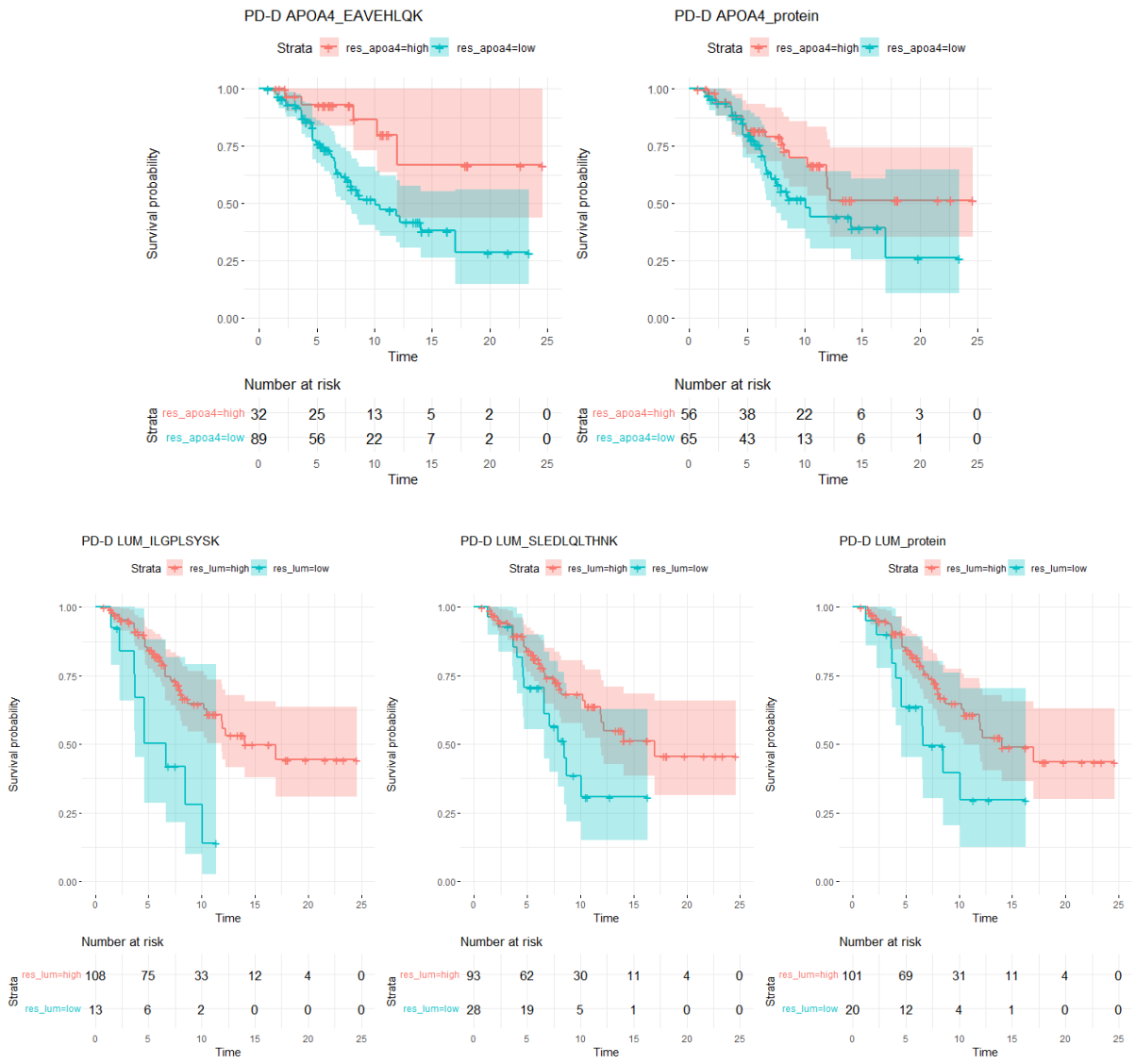

FIGURE 72 – SURVIVAL CURVES OF LUMICAN AND APOLIPOPROTEIN A-IV LEVELS AS RISK FOR DEMENTIA CONVERSION IN PARKINSON'S DISEASE

Cox regression survival curves for Lumican and Apolipoprotein A-IV, with risk for Parkinson's Disease (PD) patients to develop PD dementia equivalent Montreal Cognitive Assessment (MoCA) <21 or over time. Graphs show survival curves for both peptides and protein. High and Low protein levels set by optimal cut-off. Trendline shows mean survival probability and 95% confidence interval.

As it was noticed from the discovery study that sex had a major impact on several biomarker candidates, the PRM data was also analysed separately for males and females. For male PD patients the SAA1 peptide FFGHGAEDSLADQAANEWGR correlated with UPDRS III ($p = 0.028$), Hoehn & Yahr ($p = 0.020$), and MoCA ($p = 0.015$), and SAA1 protein levels correlated with MoCA ($p = 0.030$). This is in accordance with the discovery analysis, where SAA1 was a strong motor and cognition severity marker in male PD patients. Moreover, TGFBI protein levels were increased in male PD compared with controls ($t = 2.17$, $p = 0.033$). The APOA4 peptide EAVEHLQK was associated with conversion to PDD (HR = 0.27, $p = 0.045$), the LUM ILGPLSYSK peptide was associated with conversion to PDMCI (HR = 0.030, $p = 0.0021$). Both the LUM peptides ILGPLSYSK (HR = 0.0058, $p = 0.00017$), and SLEDLQLTHNK (HR = 0.15, $p = 0.036$) as well as protein level (HR = 0.030, $p = 0.0029$) were associated with PDD conversion in male PD. This is again similar to the discovery study where LUM predicted cognitive decline in male PD patients. For the female PD patients only the APOA4 peptide IDQTVEELR correlated with MoCA score ($p = 0.050$).

3.2.5 PRM correlation with the discovery data

Since the data was log transformed, corrected for total ion count, and adjusted for QC, the pre-processing for the PRM samples was analogous to the pre-processing of the samples in the discovery study. Therefore, the transformed PRM quantifications should hypothetically correlate well with the discovery study data. Next, it was explored how well the quantifications correlated between the PRM peptides and the protein levels from the discovery study in a correlation matrix (Figure 73). 6 of the 12 proteins from the discovery study showed a moderate correlation (Pearson $\rho > 0.25$, $p < 0.001$) with Replicates 1 and 2 in the PRM study (Figure 73). Replicate 2b (technical batch 3) showed greater variation in the correlations and was therefore excluded from downstream analysis. Replicates 1 and 2 appeared to show similar levels of correlation with the discovery data for the same peptides. The 6 proteins with the best correlations were Apolipoprotein A-IV, Apolipoprotein C-III, Complement component C9, Lumican, Vitamin K-dependent protein S, and Serum amyloid A-1 protein.

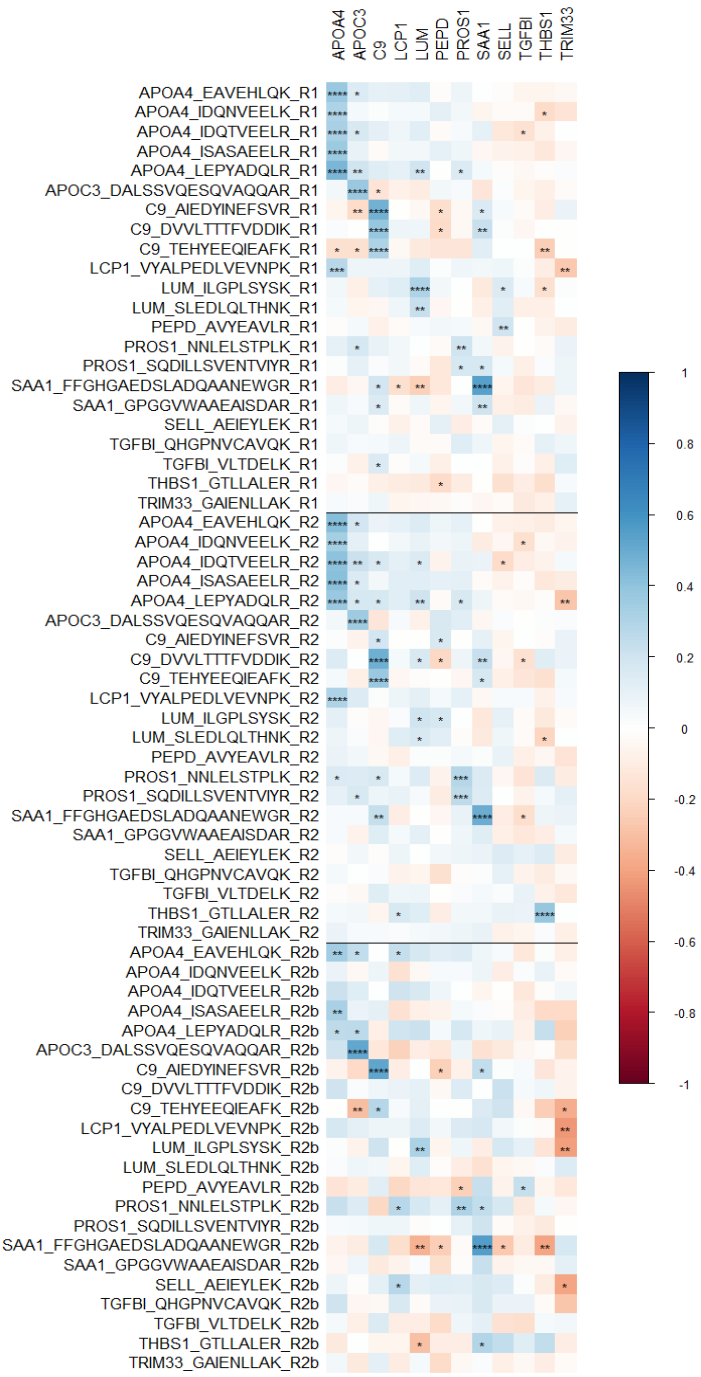

FIGURE 73 – CORRELATION PLOTS BETWEEN PROTEIN LEVELS IN THE DISCOVERY STUDY VERSUS THE VERIFICATION STUDY

*Correlation plot with peptide levels from each replicate from the verification study, versus protein levels from the discovery study. Labels as "gene_peptide_replicate". Replicates shown are replicate 1 (R1), replicate 2 (R2), and low-quality samples (R2b). Pearson correlation as colour scale, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.*

For each sample, the average of each peptide value for the two replicates was used as the average peptide quantity. The average value for all peptides quantified for a protein was used as the average protein concentration. These average peptide and protein concentrations are shown in Figure 74, along with how well they correlate with the discovery data.

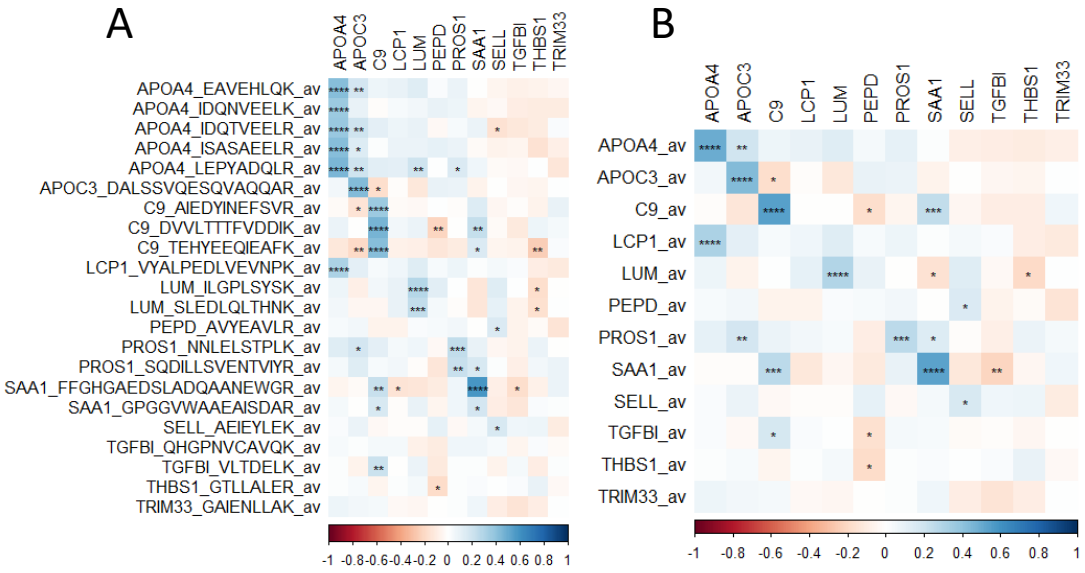

FIGURE 74 – CORRELATION PLOTS BETWEEN PROTEIN/PEPTIDE LEVELS IN THE DISCOVERY STUDY VERSUS THE VERIFICATION STUDY

A) *Correlation plot with average peptide levels from the verification study (left) versus protein levels from the discovery study (top). B) Correlation plot with average protein levels (average of peptides for each protein) from the verification study (left) versus protein levels from the discovery study (top). Pearson correlation as colour scale, *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.*

Again, it was observed that the six proteins Apolipoprotein A-IV, Apolipoprotein C-III, Complement component C9, Lumican, Vitamin K-dependent protein S, and Serum amyloid A-1 protein correlated well between the discovery data and the PRM data. Particularly 4 of the proteins, Apolipoprotein A-IV, Apolipoprotein C-III, Complement component C9, and Serum amyloid A-1 protein, displayed more pronounced correlations (Spearman ρ > 0.4). For most of these proteins, all their quantified peptides correlated similarly well with the PRM data, except for Serum amyloid A-1 (SAA1). Interestingly, for SAA1, the peptide FFGHGAEDSLADQAANEWGR correlated well with the discovery data whereas the peptide GPGGVWAAEAISDAR did not. Hence, as seen in the PRM analysis, the FFGHGAEDSLADQAANEWGR peptide correlated with MoCA and Hoehn & Yahr similar to the discovery study. Correlation statistics from the PRM protein versus discovery protein correlations are summarised in Table 27.

TABLE 27 – CORRELATIONS BETWEEN PROTEIN QUANTITIES FROM THE DISCOVERY STUDY VERSUS THE VERIFICATION STUDY

Pearson r and p-values for correlations between protein quantities from the discovery study versus the parallel reaction monitoring verification study.

Protein name	*Gene name*	*Pearson r*	*p-value*
Apolipoprotein A-IV	*APOA4*	0.49	4.17E-13
Apolipoprotein C-III	*APOC3*	0.46	1.08E-11
Complement component C9	*C9*	0.56	6.62E-18
Plastin-2	*LCP1*	0.07	0.329351
Lumican	*LUM*	0.31	1.09E-05
Xaa-Pro dipeptidase	*PEPD*	0.01	0.850986
Vitamin K-dependent protein S	*PROS1*	0.30	2.39E-05
Serum amyloid A-1 protein	*SAA1*	0.56	5.57E-17
L-selectin	*SELL*	0.18	0.013897
Transforming growth factor-beta-induced protein ig-h3	*TGFBI*	-0.04	0.60908
Thrombospondin-1	*THBS1*	0.11	0.185316
E3 ubiquitin-protein ligase TRIM33	*TRIM33*	0.01	0.950151

To summarise, 22 peptides from 12 candidate biomarker proteins were analysed on the mass spectrometer using parallel reaction monitoring (PRM). The data was adjusted for total ion cound and corrected to inter run QC samples. Peptides and proteins were analysed separately. APOC3, TGFBI and one peptide (TEHYEEQIEAFK) of C9 were changed in PD versus healthy controls. One peptide of SAA1 (FFGHGAEDSLADQAANEWGR) was associated with motor and cognitive function. LUM and APOA4 were associated with cognitive decline in PD. Several of the peptides were more significantly changed in male PD patients. Comparing the protein and peptide levels with the discovery study revealed 4 proteins (APOA4, APOC3, C9, and SAA1) correlated well.

3.3 Neurofilament light chain

Neurofilament light chain (NfL) has particularly in recent years been one the most well studied biomarkers for neurodegeneration. Although its increase in PD is generally modest [194], it has been a consistent marker for cognitive impairment and dementia in PD and appear to be linked with future decline in both disease severity [195] as well as cognition[196-198]. NfL is also elevated in DLB and atypical Parkinsonian disorders (APDs) such as MSA, PSP, and CBD. Hence it was decided to also measure plasma NfL in the cohort that was used for the mass spectrometry study. This would be used as a benchmark for a well-studied PD candidate biomarker, with has shown reproducible results for accurate diagnosis against APDs, and promising results in correlations with cognitive and motor severity.

Plasma NfL levels were quantified on the Simoa, and the data analysed and correlated with disease severity and cognitive scores. A first round of analysis found plasma NfL was elevated in PD plasma compared to healthy controls (HC), and this difference was driven by the PD with cognitive impairment (PDCI) which was defined as MoCA score <26. NfL levels in HC were comparable with PD with no dementia (PDND). Moreover, NfL levels correlated with both H&Y and MoCA scores. Plots are shown in Figure 75.

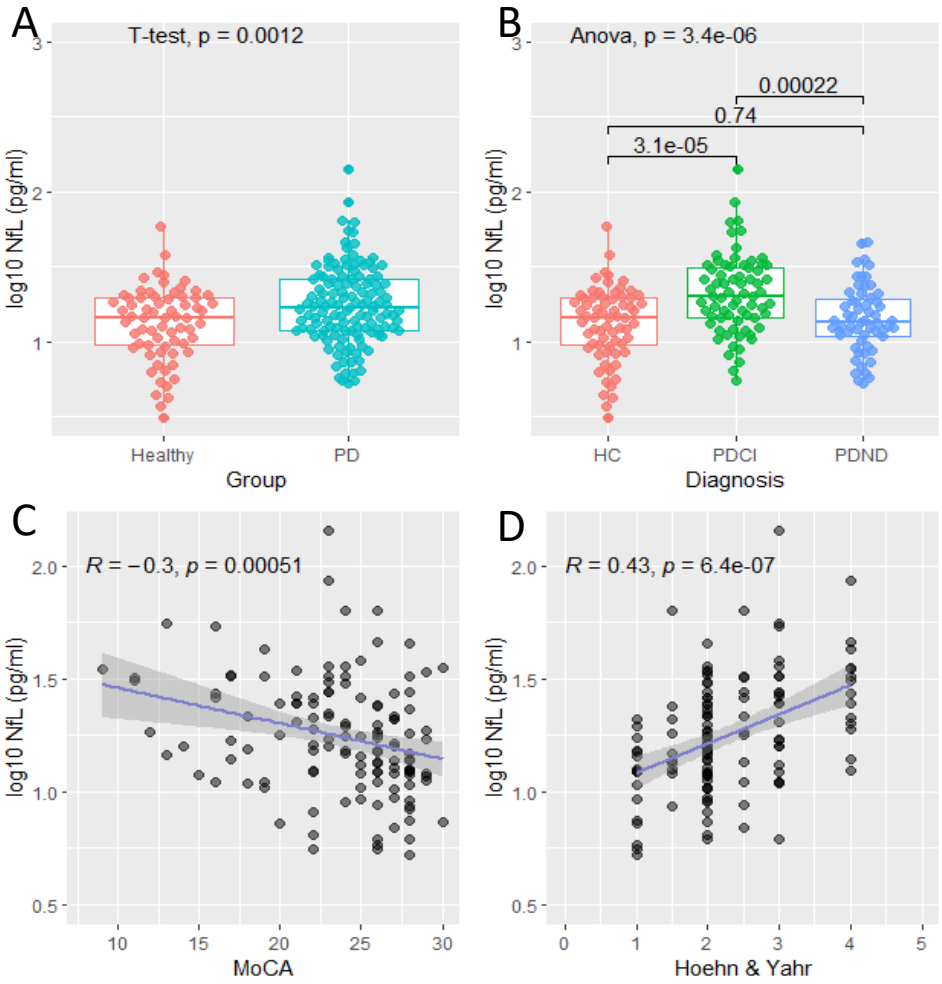

FIGURE 75 – PLASMA NEUROFILAMENT LIGHT CHAIN (NFL) IN PARKINSON'S DISEASE (PD)

A) Group level difference in plasma NfL between PD and Healthy Controls (HC) (t-test and p-value displayed). B) PD group split into PD with cognitive impairment (PDCI) where MoCA ≥26, and PD with no dementia (PDND) where MoCA >26. P-values for Anova and Tukey post hoc test shown. C) Correlation between NfL levels and MoCA score, D) correlation between.

It is known NfL levels vary greatly with age [199], hence NfL scores were adjusted for age and sex with a linear model and repeated the analysis (Figure 76). There was still a significant elevation of plasma NfL in PD patients, pronounced in the PDCI group. Moreover, the NfL levels correlated with Hoehn & Yahr, however, the correlation with MoCA score was no longer significant.

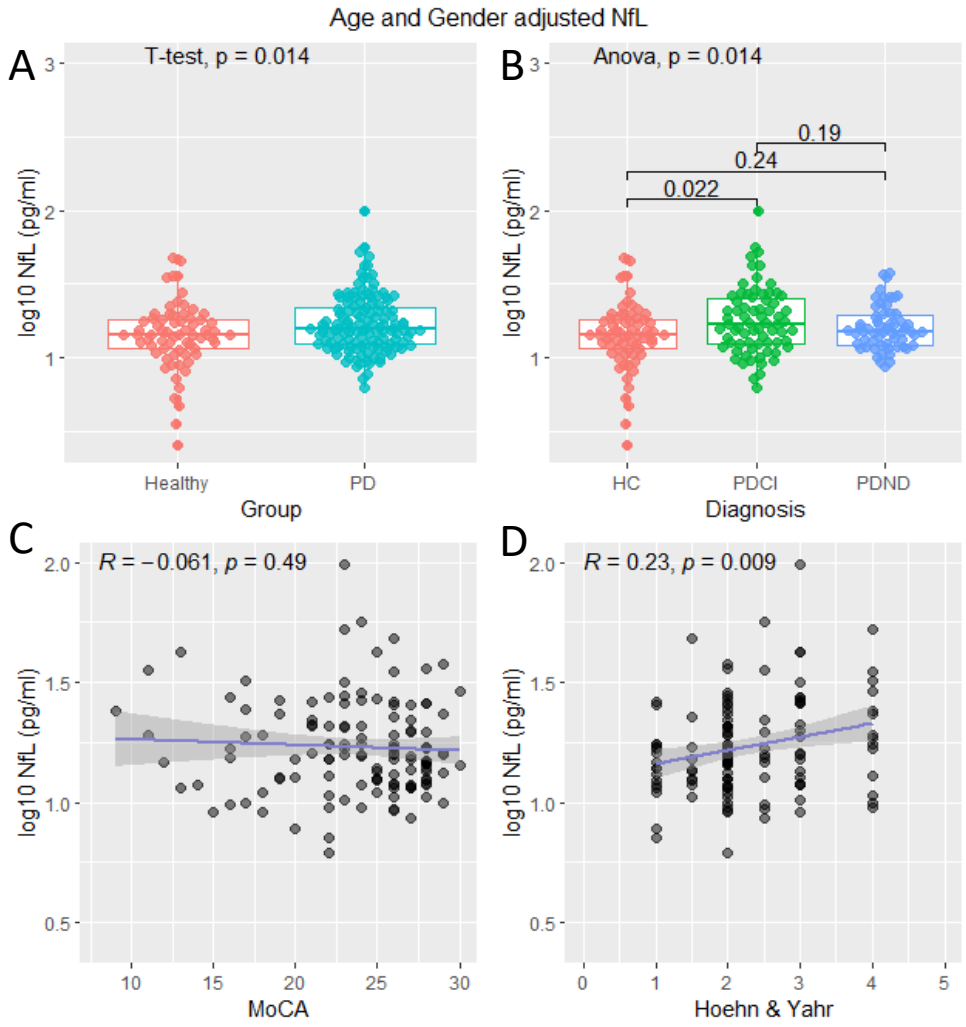

FIGURE 76 – PLASMA NEUROFILAMENT LIGHT CHAIN (NFL) IN PARKINSON'S DISEASE (PD) – AGE AND SEX ADJUSTED

A) Group level difference in plasma NfL between PD and Healthy Controls (HC) (t-test and p-value displayed). B) PD group split into PD with cognitive impairment (PDCI) where MoCA ≥26, and PD with no dementia (PDND) where MoCA >26. P-values for Anova and Tukey post hoc test shown. C) Correlation between NfL levels and MoCA score, D) correlation between.

As the association between plasma NfL levels and cognitive score in PD patients was heavily attenuated after correcting for age and sex, it was explored how NfL varied with these demographic covariates in our cohort (Figure 77), both for the entire cohort, and for HC on their own. It was observed sex had little effect on NfL levels, but NfL levels were significantly increased with age. This was true for HC as well, suggesting the increase is not only due to PD related disease progression or cognitive decline.

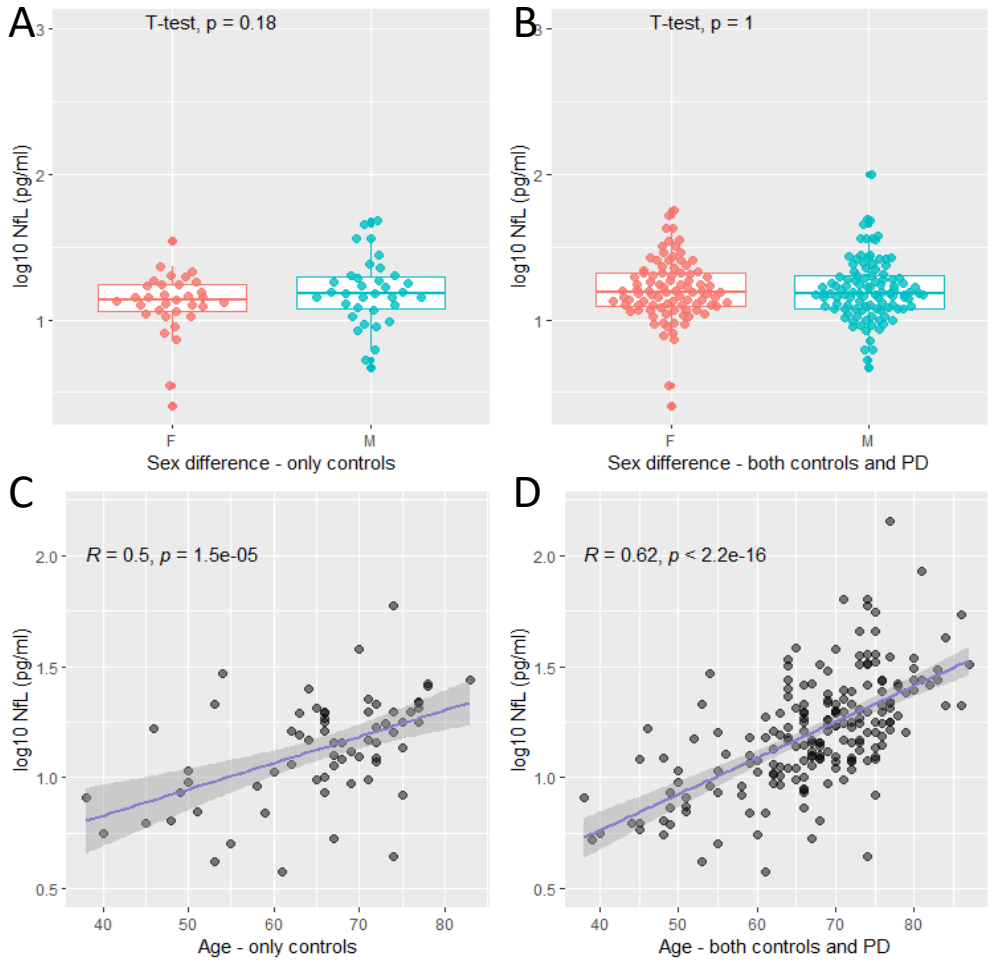

FIGURE 77 - PLASMA NEUROFILAMENT LIGHT CHAIN (NFL) ASSOCIATION WITH AGE AND SEX

A, B) Group level differences between in plasma NfL between males (M) and females (F) for healthy control subjects and whole cohort, respectively. T-test with p-value shown. C, D) Correlation between plasma NfL levels and age for healthy control subjects and whole cohort, respectively. Pearson r with p-value shown.

Next, it was investigated how well NfL would reflect cognitive decline in PD patients. A cox proportional hazard regression models was constructed for PD patients converting to PDMCI or PDD (Figure 78). Higher concentrations of NfL were associated with earlier cognitive decline. However, when the data was adjusted for age and sex, NfL was no longer associated with faster cognitive decline. The statistics are summarised in Table 28, and show age was the major predictor for cognitive decline.

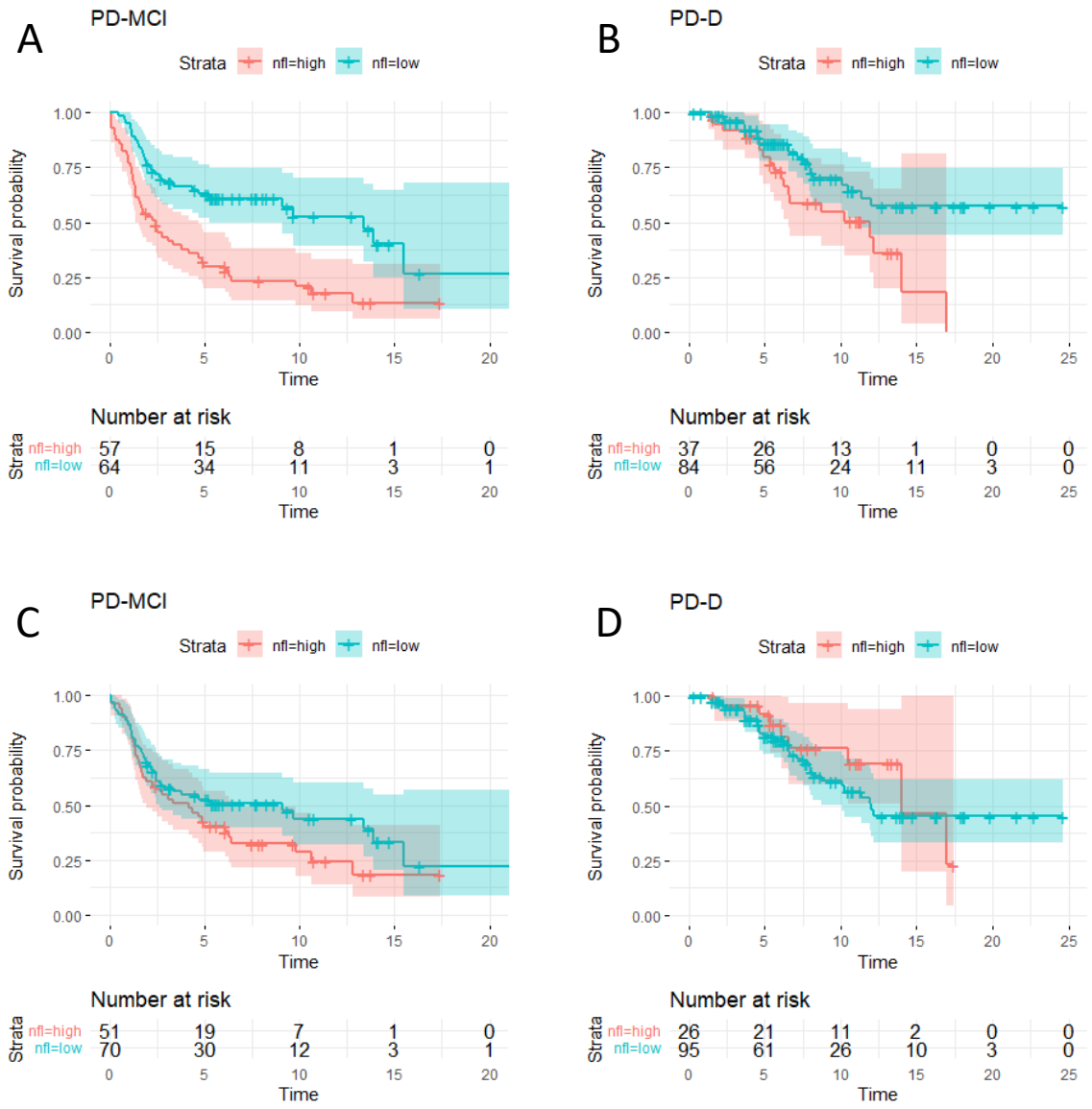

FIGURE 78 – SURVIVAL CURVES OF NEUROFILAMENT LIGHT CHAIN (NFL) LEVELS AS RISK FOR MILD COGNITIVE IMPAIRMENT AND DEMENTIA CONVERSION IN PARKINSON'S DISEASE.

Cox regression survival curves for NfL, with risk for Parkinson's Disease (PD) patients to develop A) PD mild cognitive impairment equivalent (MCI) Montreal Cognitive Assessment (MoCA) <26, and B) dementia (PDD) equivalent MoCA<21 over time. C, D Survival curves for developing PDMCI and PDD following age and sex adjusted NfL levels. High and Low protein levels set by optimal cut-off. Trendline shows mean survival probability and 95% confidence interval.

TABLE 28 – COX REGRESSION OF PARKINSON'S DISEASE DEVELOPING MILD COGNITIVE IMPAIRMENT OR DEMENTIA - NEUROFILAMENT LIGHT CHAIN (NFL)

Cox regression of NfL levels in Parkinson's Disease patients versus risk of declining cognitively to Montreal Cognitive Assessment <26 and <21, with age and sex as covariates. Columns show protein gene name, Cox coefficient, Hazard ratio, standard error, and p-value.

	PDMCI (MoCA < 26) Conversion				*PDD (MoCA < 21) Conversion*			
Variables	*coef*	*HR exp(coef)*	*se(coef)*	*p-value*	*coef*	*HR exp(coef)*	*se(coef)*	*p-value*
log NFL conc.	0.122	1.130	0.605	0.84	0.101	1.106	0.819	0.902
Age	0.126	1.134	0.020	**7.92E-10**	0.099	1.105	0.024	**4.86E-05**
Sex Male	-0.003	0.997	0.240	0.991	0.205	1.227	0.340	0.547

Following this, the NfL biomarker performance in PD was compared against our new candidate biomarkers from the mass spectrometry study. After accounting for age and sex, NfL was a better diagnostic PD biomarker than most verified peptides, and comparable to TGFBI and the TEHYEEQIEAFK peptide of C9. The SAA1 peptide FFGHGAEDSLADQAANEWGR correlated better with MoCA than NfL, but NfL correlated better with Hoehn & Yahr. Both LUM and APOA4 were better predictors of cognitive decline in PD than NfL. However, many candidate biomarkers from the discovery study that were not verified using the PRM technique have the potential to outperform NfL if validated in a different way, and many of those candidate proteins do not seem to covary with age.

NfL is often described as a sensitive but unspecific biomarker that reflects axonal damage and level of neurodegeneration [92]. NfL levels were therefore correlated with the protein levels from the mass spectrometry study, as that could give some insight in which proteins reflect the axonal damage/neurodegenerative component of PD. Most of these proteins (Table 29) also correlated with age, and it was difficult to determine whether both proteins correlated with age, or if they correlated with each other. A few of these proteins did however correlate with NfL without correlating with age and are highlighted in Table 29.

TABLE 29 – CORRELATIONS BETWEEN PLASMA NFL LEVELS AND PROTEINS FROM THE DISCOVERY STUDY

Correlations between plasma levels of proteins from the mass spectrometry discovery study, and their correlations with plasma Neurofilament Light Chain (NfL) levels and age. Columns show protein and gene name, Pearson r and p-value for NfL and Age. Proteins not correlating with age are highlighted.

Protein	*Gene*	*r NfL*	*p NfL*	*r Age*	*p Age*
EGF-containing fibulin-like extracellular matrix protein 1	EFEMP1	0.42	4.59E-10	0.39	1.16E-08
Fibulin-5	FBLN5	0.42	6.52E-06	0.30	1.66E-03
m-AAA protease-interacting protein 1, mitochondrial	MAIP1	0.35	2.75E-06	0.20	7.35E-03
Complement component C9	C9	0.35	4.94E-07	0.33	2.66E-06
V-type proton ATPase 116 kDa subunit a isoform 4	ATP6V0A4	0.34	3.69E-04	0.20	4.29E-02
Cystatin-C	CST3	0.33	4.31E-06	0.28	1.07E-04
Proline-rich protein 36	PRR36	0.33	2.94E-06	0.19	7.37E-03
Fibrinogen-like protein 1	***FGL1***	***0.31***	***1.57E-04***	***0.11***	***1.78E-01***
Inter-alpha-trypsin inhibitor heavy chain H3	ITIH3	0.31	8.87E-06	0.36	1.91E-07
Serine/threonine-protein kinase N2	PKN2	0.31	8.69E-04	0.20	3.02E-02
SRC kinase signaling inhibitor 1	***SRCIN1***	***0.30***	***1.73E-03***	***0.05***	***6.33E-01***
Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta	***PIK3C2B***	***0.30***	***1.64E-03***	***0.16***	***1.10E-01***
Pyruvate dehydrogenase protein X component, mitochondrial	PDHX	0.30	1.66E-03	0.31	1.10E-03
Complement factor D	CFD	0.29	3.51E-04	0.35	8.48E-06
Coiled-coil domain-containing protein 18	CCDC18	0.28	1.19E-03	0.21	1.70E-02
Contactin-6	***CNTN6***	***0.26***	***6.47E-03***	***0.06***	***5.21E-01***
Peptidase inhibitor 16	PI16	0.26	2.36E-04	0.16	2.21E-02
N-acetyl-B-glucosaminyl-glycoprotein 4-B-N-acetylgalactosaminyltransferase 1	B4GALNT4	0.26	2.56E-03	0.24	5.09E-03
Prostaglandin-H2 D-isomerase	PTGDS	0.26	3.60E-04	0.22	2.32E-03
Fibulin-1	FBLN1	0.25	3.43E-04	0.16	2.36E-02
PDZ domain-containing protein 4	***PDZD4***	***0.25***	***3.88E-03***	***0.14***	***9.57E-02***
Beta-2-microglobulin	B2M	0.25	1.08E-03	0.24	1.41E-03
Neuronal tyrosine-phosphorylated phosphoinositide-3-kinase adapter 2	***NYAP2***	***0.24***	***9.23E-03***	***0.07***	***4.60E-01***
Protein mono-ADP-ribosyltransferase PARP15	PARP15	0.24	2.88E-03	0.24	2.64E-03
Integral membrane protein GPR180	GPR180	0.23	4.77E-03	0.17	4.39E-02

To summarize, the plasma NfL levels in our cohort showed a strong positive correlation with age, both for PD patients and healthy controls. After correcting for age and sex, NfL was still elevated in PD, and was pronounced in the cognitively impaired patients. Moreover, NfL correlated with H&Y disease severity. Overall, NfL showed modest biomarker performance in our cohort. Many of the generated biomarker candidates from the discovery study, and some of the verified proteins, would potentially be better biomarkers for PD and associated cognitive and motor changes.

3.5 Complement study

3.5.1 Study overview

Proteins related to the complement system have emerged as candidate plasma biomarkers in several studies, in fact about 20% of the reproduced biomarker candidates in PD plasma proteomic studies are involved in the complement system [114]. Additionally, the complement and coagulation cascade was the most implicated pathway in PD, and C9 was one of the top biomarker candidates. This was enough evidence to warrant a validation study of complement factors in PD plasma. Two commercially available Luminex® multiplex complement panels and a separate C9 ELISA were run. PD patients were compared with healthy controls (HC), and the atypical parkinsonian disorders Corticobasal Syndrome (CBS) and Progressive Supranuclear Palsy (PSP). PSP and CBS share both some symptomatology and pathology [200] and were pooled into one 4-repeat (4R)-Tauopathy group as a neurological control. Finally, a CH50 assay was performed to assess overall classical pathway complement activity in PD serum. An overview of the complement system and its different activity pathways are shown in Figure 79.

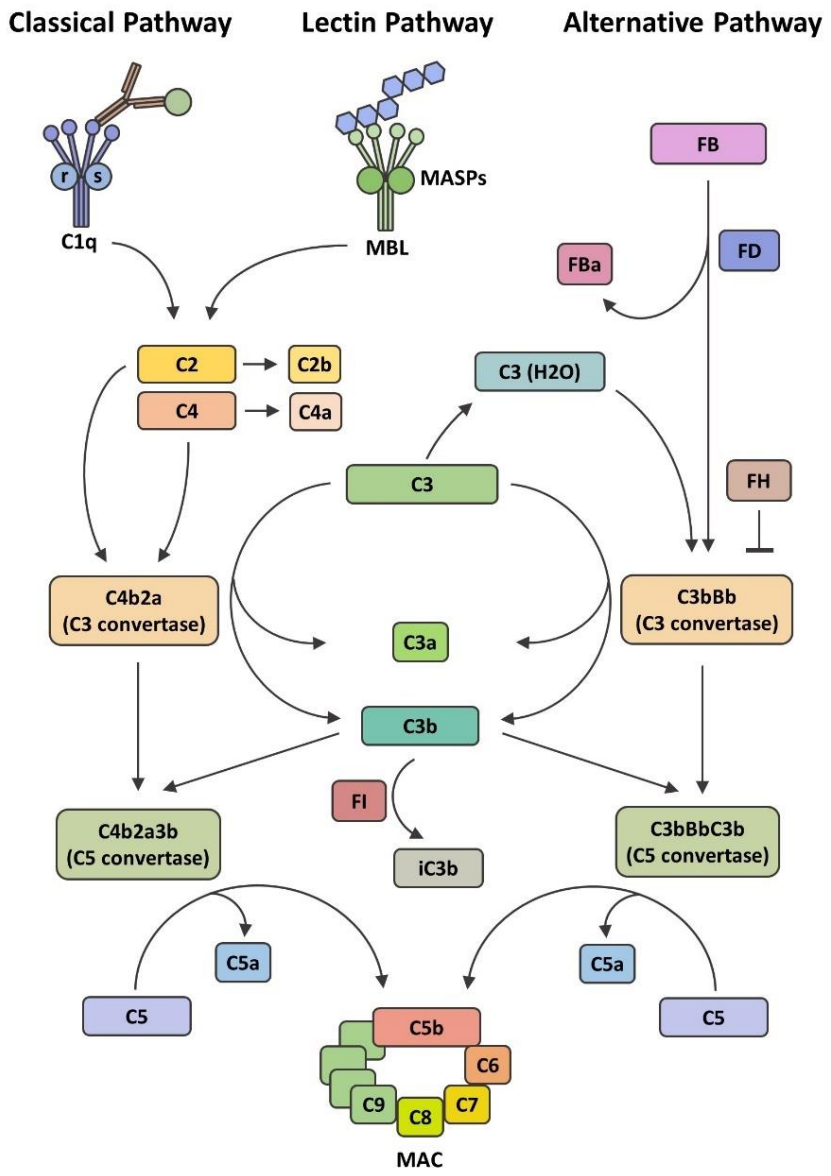

FIGURE 79 – SCHEMATIC OF THE COMPLEMENT SYSTEM

Overview of the complement system including all the complement proteins measured. Abbreviations: MBL Mannose Binding Lectin, MASPs MBL Associated Serine Proteases, MAC Membrane Attack Complex, FB Factor B, FD Factor D, FH Factor H.

3.5.2 Plasma levels of complement factors in PD and APD

Plasma concentrations were quantified for 12 complement proteins in HC, PD and 4R-Tauopathies using immunoassays (summarised in Table 30). The concentrations of complement proteins were log-transformed and corrected for age and sex before analysis. Plasma levels of C1q and C3 were significantly different between investigated groups (ANOVA p=0.0041 and p=0.0057 respectively;

Figure 80). Tukey's post hoc tests revealed lower concentrations in 4R-Tauopathies compared to both PD and HC, for both C1q (p=0.0029, p=0.025 respectively; Figure 80) and C3 (p=0.0095, p=0.0068 respectively; Figure 80). Receiver operating characteristic (ROC) curves were carried out and area under curve (AUC) was calculated to assess how well C1q and C3 would perform as biomarkers. The AUCs for C1q were 0.73 and 0.70 for 4R-Tauopathies versus PD and HC respectively, and for C3 AUCs were 0.67 and 0.69 for 4R-Tauopathies versus PD and HC, respectively.

There were no significant differences between groups for any of the other analytes. No differences were found between PD patients and HC, for any of the factors. Sex-specific analysis, comparing PD versus HC for males and females separately did not show any differences either (data not shown).

Next, ratios between two of the complement factors (C4 and C5) and their respective cleavage products (C4b and C5a) were calculated, as a proxy to measure increased complement cleavage and activity. No differences were found between the groups (data not shown).

PCA plots were constructed for all the complement proteins to visualise the variation in the data (Figure 81). The plot revealed 38.5% of the variation was explained on PC1, with covariance between most of the complement factors, and large overlap between the groups when plotting PC1 and PC2.

Taken together this shows that C1q and C3 levels were decreased in 4R-Tauopathies plasma compared to both PD and HC, and that there were no differences in complement factor concentrations between PD and HC.

TABLE 30 – DEMOGRAPHICS OF THE BIOPARK AND AETIONOMY COHORTS, CLINICAL ASSESSMENTS, COMPLEMENT PROTEIN QUANTIFICATIONS AND CH50.

All clinical data presented as median (range), complement concentrations and CH50% as mean (standard deviation). Age and disease duration in years, Movement Disorder Society Unified Parkinson's Disease Rating Scale part 1-4 (MDS UPDRS I,II,III,IV), Hoehn & Yahr, Montreal Cognitive Assessment (MoCA), Beck Depression Inventory-2 (BDI II), Montgomery-Åsberg Depression Rating Scale (MADRS), Mental Fatigue Scale (MFS), Hospital Anxiety and Depression Scale (HADS) depression and anxiety sub scores, Non-Motor Symptom Questionnaire (NMSQ), The Parkinson's Disease Questionnaire (PDQ 39), Pittsburgh Sleep Quality Index (PSQI), Levodopa Equivalent Daily Dose (LEDD). Plasma levels for complement factors C3, C4, C4b, C5, C5a, C9, factors B (FB), D (FD), H (FH), I (FI), and Mannose Binding Lectin (MBL) presented as log10 concentrations corrected for age and sex. CH50 % serum as percentage serum for 50% haemolysis. P-values represent chi-square for male to female comparisons, Wilcoxon for age, ANOVA for complement concentrations, 2-tailed t-test for CH50.

	BIOPARK				*AETIONOMY*		
	Parkinson's Disease	***Healthy Controls***	***4R-Tauopathies***	***p-value***	***Parkinson's Disease***	***Healthy Controls***	***p-value***
n (Female : Male)	81 (31 : 50)	48 (28 : 20)	23 (14 : 9)	0.035	58 (16 : 42)	20 (13 : 7)	0.0028
Age	69 (49 - 84)	67 (50 - 83)	73 (66 - 96)	0.0046	64 (41 - 75)	65 (52 - 97)	0.99
Disease Duration	3.31 (0 - 23.46)	-	1.52 (0- 8.13)	-	2.98 (0 - 10.82)	-	-
MDS UPDRS I	-	-	-	-	9 (2 - 26)	-	-
MDS UPDRS II	-	-	-	-	11 (1 - 23)	-	-
MDS UPDRS III	31 (3 - 64)	-	-	-	24 (3 - 51)	-	-
MDS UPDRS IV	-	-	-	-	1 (0 - 10)	-	-
Hoehn & Yahr	2 (0 - 4)	-	-	-	2 (1 - 3)	-	-
MoCA	25 (12 - 30)	-	-	-	26 (18 - 35)	-	-
BDI II	10 (0 - 43)	-	-	-	10 (0 - 38)	-	-
MADRS	6 (0 - 32)	-	-	-	8 (1 - 36)	-	-
MFS	10 (0 - 24)	-	-	-	9 (1 - 32)	-	-
HADS Anxiety	5 (0 - 13)	-	-	-	5 (0 - 19)	-	-
HADS Depression	3 (0 - 14)	-	-	-	2.5 (0 - 17)	-	-
NMSQ	9 (0 - 20)	-	-	-	9 (2 - 20)	-	-
PDQ 39	21.6 (0 - 60.4)	-	-	-	19.7 (3.4- 71.8)	-	-
PSQI	7 (0 - 18)	-	-	-	7 (3 - 18)	-	-
LEDD	500 (0 - 2235)	-	-	-	445 (0 - 1606)	-	-
C1q (µg/ml)	1.90 (0.15)	1.88 (0.18)	1.77 (0.17)	0.0041	-	-	-
C3 (µg/ml)	1.76 (0.29)	1.79 (0.32)	1.54 (0.33)	0.0057	-	-	-
C4 (µg/ml)	2.14 (0.18)	2.15 (0.23)	2.10 (0.22)	0.73	-	-	-
FB (µg/ml)	2.14 (0.15)	2.14 (0.18)	2.12 (0.20)	0.84	-	-	-
FH (µg/ml)	2.34 (0.13)	2.34 (0.16)	2.30 (0.15)	0.45	-	-	-
C4b (µg/ml)	1.07 (0.12)	1.08 (0.13)	1.05 (0.09)	0.71	-	-	-
C5 (µg/ml)	1.12 (0.09)	1.11 (0.10)	1.13 (0.11)	0.78	-	-	-
C5a (µg/ml)	-3.15 (0.11)	-3.15 (0.09)	-3.14 (0.09)	0.90	-	-	-
FD (µg/ml)	0.49 (0.15)	0.54 (0.17)	0.49 (0.12)	0.26	-	-	-
MBL (µg/ml)	0.05 (0.53)	0.10 (0.50)	0.07 (0.50)	0.82	-	-	-
FI (µg/ml)	1.33 (0.12)	1.37 (0.13)	1.36 (0.10)	0.32	-	-	-
C9 (µg/ml)	1.36 (0.12)	1.37 (0.13)	1.32 (0.12)	0.24	-	-	-
CH50	-	-	-	-	79.4 (13)	79.2 (14)	0.95

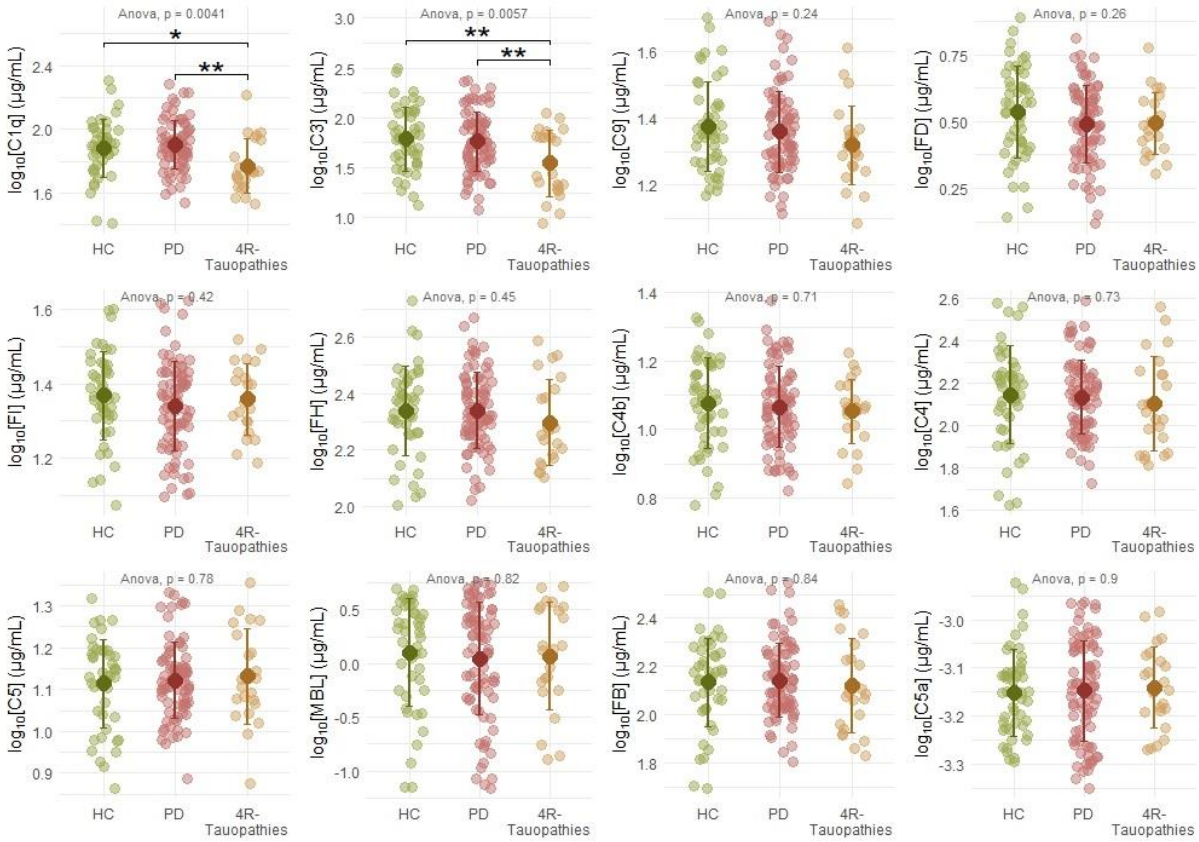

FIGURE 80 – PLASMA CONCENTRATION OF COMPLEMENT PROTEINS IN HEALTHY CONTROLS (HC), PARKINSON'S DISEASE (PD), AND FOUR-REPEAT TAUOPATHIES (4R-TAUOPATHIES).

*Plasma levels for complement factors C3, C4, C4b, C5, C5a, C9, factors B (FB), D (FD), H (FH), I (FI), and Mannose Binding Lectin (MBL) in three cohort groups (HC, PD, 4R Tauopathies) plotted in order of significance (ANOVA with Tukey's post hoc test for multiple comparisons, *p < 0.05, **p < 0.01). Plots showing individual data points, mean and standard deviation.*

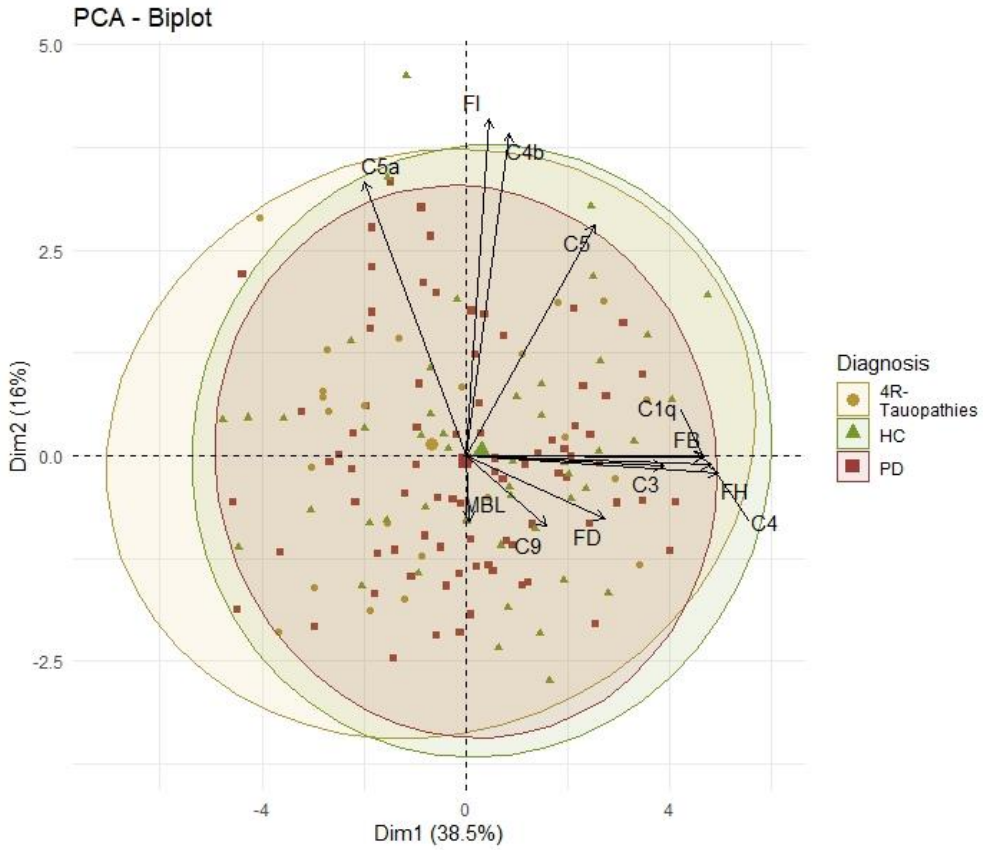

FIGURE 81 – PCA FOR COMPLEMENT FACTORS AND DIAGNOSES

Principle component analysis (PCA) of plasma concentrations for 12 complement factors C3, C4, C4b, C5, C5a, factors B (FB), D (FD), H (FH), I (FI), and Mannose Binding Lectin (MBL) in Parkinson's disease (PD) adjusted for age and sex. Ellipses represent 95% confidence interval for Parkinson's disease (PD), Four repeat tauopathies (4R Tauopathies) and Healthy controls (HC).

3.5.3 Association between complement factors and clinical assessments

Next, correlations were performed between complement factors and the clinical scores within the PD group. Scales for disease stage, motor symptoms and several non-motor symptoms were selected. The levels of complement proteins were log-transformed and corrected for age, sex, and disease duration for all PD patients, and for age and disease duration only in sex-specific analysis, before correlating with several clinical parameters using Spearman's ρ.

Plasma C1q correlated negatively with cognitive performance (MoCA) (p=0.041; Figure 82), FD correlated negatively with the mental fatigue (MFS) (p=0.043; Figure 82), and FI correlated negatively with motor severity (MDS UPDRS III) (p=0.024; Figure 82). Sex-specific correlations were performed to explore complement differences between male and female PD patients. For female patients, several of the complement factors correlated positively with NMS. Multiple clinical parameters correlated

positively with C3, including HADS Anxiety (p=0.0073), Depression (p=0.012), NMSQ (p=0.0015), and BDI-II (p=0.038). Additionally, HADS Anxiety correlated with positively with several other complement factors, including C1q (p=0.0056), FH (p=0.035), and C5 (p=0.017). Moreover, non-motor symptom severity (NMSQ) correlated positively with C5 (p=0.044), PDQ 39 with C5a (p=0.020), and Hoehn & Yahr disease stage correlated negatively with C4b (p=0.041). As for the male patients, C3 was negatively correlated with PDQ 39 (p=0.016), FD was negatively correlated with MFS (0.040), MBL positively correlated with MoCA (p=0.023), and FI positively correlated with MDS UPDRS III (p=0.023). Interestingly, NMS severity in PD patients were generally negatively correlated with plasma complement levels, which is the opposite of what was observed in female patients.

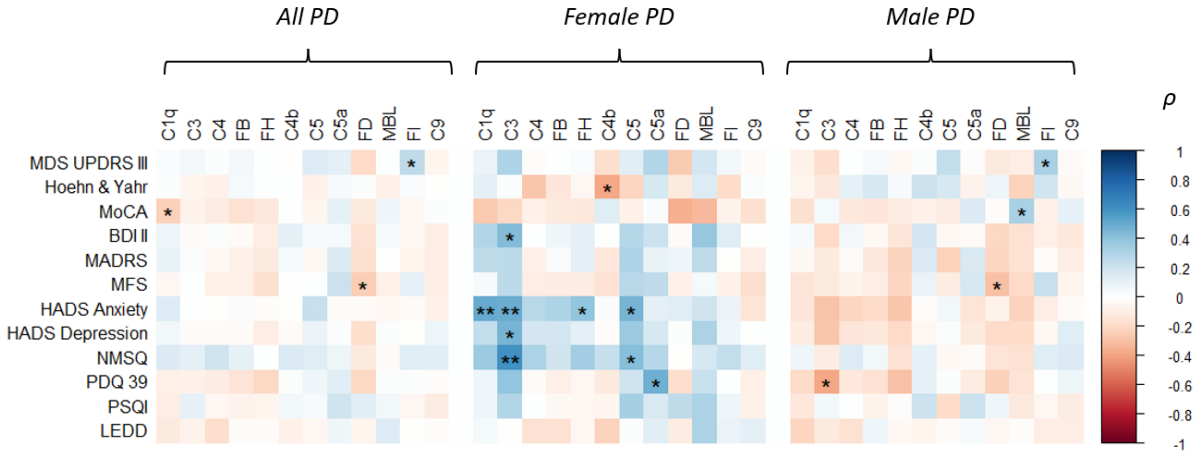

FIGURE 82 – CORRELATION MATRIX BETWEEN CLINICAL SCORES AND COMPLEMENT FACTORS

*Plasma levels for complement factors C3, C4, C4b, C5, C5a, factors B (FB), D (FD), H (FH), I (FI), and Mannose Binding Lectin (MBL) in Parkinson's disease (PD) adjusted for age and disease duration correlated with clinical scales, both using all PD patients combined, and split into male and female patients (Spearman's ρ, *p<0.05, **p<0.01). Clinical scales include Movement Disorder Society Unified Parkinson's Disease Rating Scale part 3 (MDS UPDRS III), Hoehn & Yahr, Montreal Cognitive Assessment (MoCA), Beck Depression Inventory-2 (BDI II), Montgomery-Åsberg Depression Rating Scale (MADRS), Mental Fatigue Scale (MFS), Hospital Anxiety and Depression Scale (HADS) depression and anxiety sub scores, Non-Motor Symptom Questionnaire (NMSQ), The Parkinson's Disease Questionnaire (PDQ 39), Pittsburgh Sleep Quality Index (PSQI), Levodopa Equivalent Daily Dose (LEDD).*

3.5.4 Classical pathway activity in PD sera

Although no differences were observed in the assayed plasma complement proteins between PD and HC, it was further investigated whether overall complement activity, and indirectly complement function, was altered peripherally in PD. Hence, using serum from a smaller separate cohort of PD and HC where classical pathway activity was measured. There was no significant difference in CH50 between PD and HC (2-tailed T-test p=0.95, Figure 83A). Positive correlations between CH50 and mental fatigue (MFS) was observed for PD patients p<0.05; Figure 83B). Again, sex-specific analyses were performed, and it was observed the correlation between MFS and CH50 was particularly driven by male PD patients.

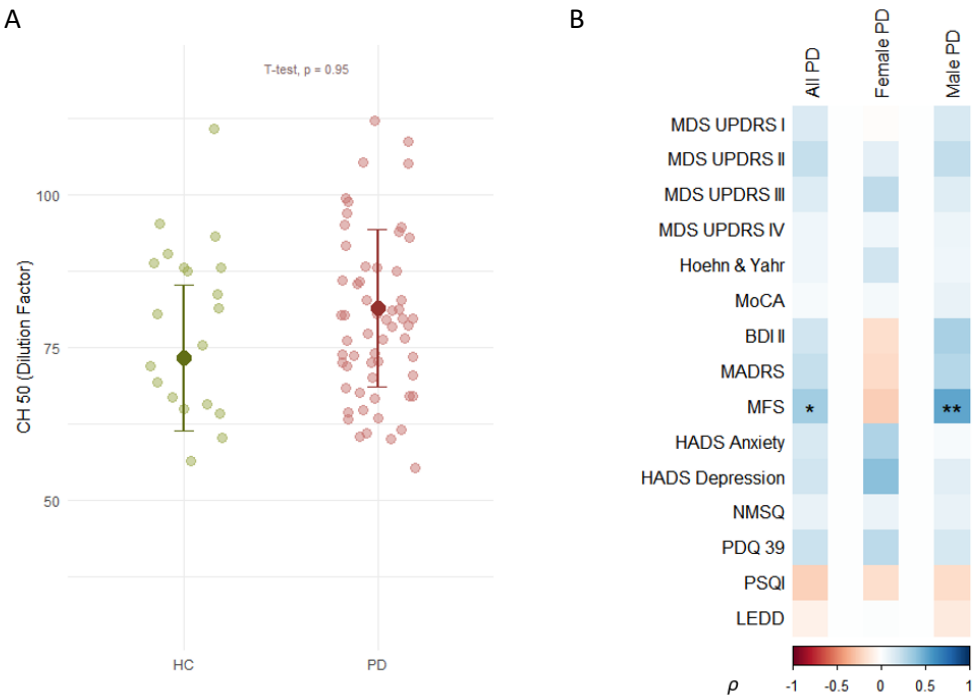

FIGURE 83 – CH50 ASSAY IN PARKINSON'S DISEASE (PD) AND HEALTHY CONTROLS (HC), AND FOR CLINICAL SCORES IN PD

*A) Dilution factor of serum for 50% haemolysis in the CH50 assay for PD versus HC show no difference in complement activity (2-tailed T-test). B) the CH50 score is correlated with both using all PD patients combined and split into male and female patients (Spearman's ρ, *p<0.05, **p<0.01). Clinical scales include disease duration, Movement Disorder Society Unified Parkinson's Disease Rating Scale part 1-4 (MDS UPDRS I,II,III,IV), Hoehn & Yahr, Montreal Cognitive Assessment (MoCA), Beck Depression Inventory-2 (BDI II), Montgomery-Åsberg Depression Rating Scale (MADRS), Mental Fatigue Scale (MFS), Hospital Anxiety and Depression Scale (HADS) depression and anxiety sub scores, Non-Motor Symptom Questionnaire (NMSQ), The Parkinson's Disease Questionnaire (PDQ 39), Pittsburgh Sleep Quality Index (PSQI), Levodopa Equivalent Daily Dose (LEDD).*

To summarise, as both our study, as well as most other proteomic studies on PD plasma, discovered alterations in complement proteins, we decided to attempt to validate them. 12 complement proteins were quantified in PD, HC, and 4R-Taupathy plasma.

Both plasma C1q and C3 were decreased in 4R-Taupathies compared to both HC and PD. However no group level differences were observed between PD and HC for any of the complement factors. A few of the complement factors correlated with fatigue, cognition, and motor symptoms in PD patients. Interestingly, C1q that has been implicated in neurodegeneration [201], was increased in PD patients of both sexes with lower MoCA scores. Additionally, C3 correlated with several non-motor scores in female PD patients. We further used the CH50 assay to quantify complement activity in PD and HC sera. Again, no group difference was found between PD and HC, but increased CH50 was found male patients with mental fatigue, suggesting they had higher level of complement activity.

4. Discussion

4.1 Summary of findings

It was demonstrated in this thesis that untargeted mass spectrometry is a viable approach for discovering novel PD plasma biomarker candidates. Several biomarker candidates were discovered, some of which were successfully verified with PRM. Additionally, some candidate biomarkers showed potential for superior biomarker performance to NfL. Although some of these results are encouraging for the future of PD biomarkers, there are still many limitations with the methodologies, and future validation studies of the candidate biomarkers are essential.

4.2 Recap of the approach

The main aim of this thesis was to discover and verify novel plasma biomarkers for Parkinson's disease diagnosis and associated cognitive impairment, using a pipeline with stepwise reduction of candidate proteins (Figure 3). Biomarkers are lacking in PD, and many promising biomarkers identified in other neurological diseases have failed to be replicated in PD [94]. A hypothesis driven approach has yielded some promising findings in α-syn markers in CSF [70, 202] but has otherwise not resulted in any major breakthroughs. A few attempts have been made with an unbiased approach in plasma, usually through mass spectrometry, but the findings have been hard to replicate [114]. Reasons include underpowered studies, and processing methods that yield only high abundant plasma proteins, several orders of magnitude above the kind of markers that would reflect neurodegeneration in for example AD. Moreover, the gap between a discovery study and a validation study in a separate cohort using an antibody-based approach is often very large, which can lead to high attrition of candidate markers, which can be discouraging [133]. To tackle this, a large-scale proteomic discovery and verification study using plasma from 130 PD patients and 68 neurologically healthy individuals was attempted. Plasma samples from these individuals were processed to maximise resolution and quantification of low abundant plasma proteins accurately, while still providing a feasible protocol. Moreover, the data was matched at a 5% FDR to achieve much higher number of candidate markers (albeit with the risk of more false positives). Peptides from candidate biomarkers that could be confidently quantified in neat plasma were then validated on the mass spectrometer with PRM. This would give confidence that our results were indeed true and should be further validated in other cohorts or with other approaches. Finally, as both our results as well as previous studies had found consistent alteration in complement proteins, an independent immunoassay-based study of complement proteins was performed to validate these findings.

4.3 Top candidate biomarkers

The discovery study generated many candidate biomarkers for PD diagnosis, disease severity, cognitive severity, and cognitive decline. Although most of these proteins still need validation, many the most significant candidates form the analyses are likely to be involved in the disease pathology.

The top three most significant differentially expressed proteins for PD diagnosis were Myosin light chain kinase, smooth muscle (MYLK), Xaa-Pro dipeptidase (PEPD), and Proline-rich protein 36 (PRR36). MYLK plays a central role in smooth muscle contraction and is necessary for gut motility [203]. This is interesting as one of the earliest symptoms of some PD patients is gut dysmotility [204], and if MYLK is validated, and potentially linked with PD gut dysfunction, it could possibly be used as an early diagnostic biomarker. PEPD, a cytosolic dipeptidase also known as prolidase, was decreased in PD plasma in our study. One study found decreased plasma prolidase activity in PD patients with a concomitant increase in oxidative stress [205]. The function of PRR36 is unknown, but it is highly expressed in the cortex and pituitary gland according to the Human Protein Atlas [206]. Levels of Enhancer of polycomb homolog 2 (EPC2), Gamma-glutamyl hydrolase (GGH), and Cytospin-B (SPECC1) showed the largest effect sizes in PD compared with controls. EPC2 is not well studied, and believed to play a role in DNA repair [207]. GGH, a metabolic enzyme highly expressed in liver, gut, and kidneys [207], was found to be a promising candidate marker for PD in a proteomic study of the substantia nigra, and found to be expressed in dopaminergic neurons [208]. Interestingly, the same article also verified Nebulette as a top candidate marker that was expressed in dopaminergic neurons. Nebulette was incidentally one of the most significant proteins in our study correlating with PD cognition. SPECC1 is highly expressed in testis and brain, particularly in oligodendrocytes [206], but its function in the CNS is poorly studied.

Serum Amyloid A1 (SAA1), IQ motif containing C (IQCC), and MAP10 were all correlating strongly with MoCA score. SAA1 is an acute phase protein, and is associated with amyloidosis [209]. One recent biomarker study found elevated Serum Amyloid A in PD CSF predicted cognitive decline in PD [210]. An animal model of AD showed that overexpressed SAA1 aggravated amyloid plaque aggregation and cognitive impairment [211]. IQCC is expressed mostly in testis and neurons, and its function is not clear [206]. MAP10 has not specifically been studied in cognition, but plays a role in cell division and microtubule stability like other Microtubule Associated Proteins (MAPs) including MAP tau (MAPT) [212].

Vitamin K-dependent protein S (PROS1), Telomeric repeat-binding factor 2 (TERF2), and Obscurin (OBSCN) were the proteins that were most associated with earlier cognitive decline in PD. PROS1 primarily acts as an anti-coagulation cofactor produced by the liver but is also reported to play a key

regulatory function in hippocampal neural stem cell proliferation [213]. TERF2 plays a central role in telomere maintenance but is not well studied in neurodegeneration. However, telomeres have been linked to neurodegeneration and dementia, and a meta-analysis showed patients with AD had overall shorter telomeres [214]. OBSCN is highly expressed in skeletal muscle and involved in myofibrillogenesis but has not been studied in relation to neurodegeneration [215].

A recent review from Chelliah et al. [114] summarised the results from all plasma proteomic PD studies to date (n = 12) and identified 23 proteins that were differentially expressed in at least 2 studies [114]. Two of the proteins, APOC3 and CP were also differentially expressed in our study. Interestingly, many of the 23 proteins correlated with MoCA in our study, including SAA1, TTR, APOA4, and APOA1. This shows there is some level of concordance between our results, and previous proteomic studies in plasma.

4.4 Utility of verified biomarkers

Most of the candidate biomarker top hits mentioned above, except SAA1 and PROS1, did not have specific and reproducible peptides available to verify with PRM in plasma. Out of the top 70 proteins reflecting PD diagnosis and cognition, 12 had high quality unique peptides for a PRM verification, 6 of which (Apolipoprotein A-IV (APOA4), Apolipoprotein C-III (APOC3), Complement component C9 (C9), Lumican (LUM), Vitamin K-dependent protein S (PROS1), and Serum amyloid A-1 protein (SAA1) correlated with the quantifications from the discovery study. It was not surprising relatively few proteins had peptides that could be consistently quantified in unfractionated plasma, as few of the proteins in the discovery had peptides that were detected in most samples, despite several pre-processing steps and a 5%FDR database matching. APOC3, TGFBI, and one of the peptides from C9 were successfully verified and significantly altered in PD after correcting for age and sex, although the statistical differences were modest. Although APOC3 specifically has not been extensively studied in PD, other apolipoproteins are implicated in PD. Plasma APOA1 is generally downregulated in PD, high APOD neurons are more resistant to PD pathology, and APOE and APOJ genes are both linked to PD dementia [216]. Moreover, APOE is shown to be enriched in PD dopaminergic neurons, elevated in PD CSF, and interact with α-syn [217]. SAA1, particularly the FFGHGAEDSLADQAANEWGR peptide, was the only candidate marker that was verified and correlated significantly with both cognition and motor severity after correcting for covariates. It is a bit unclear whether it mainly reflects processes associated with motor or cognitive symptoms primarily, but it does seem to increase with disease progression independently of age. LUM and APOA4 reflected conversion to PDD in the verification study. LUM was particularly associated with cognitive decline in male PD patients. Limited number of studies are available for the role of APOA4 in PD cognition, but as discussed above, apolipoproteins

are in general implicated in PD. Lumican is a structural protein, and has previously appeared as a proteomic biomarker candidate in cognitive impairment [218], but studies on its role in PD are lacking.

Robust peptides from APOC3, C9, TGFBI, SAA1, APOA4, and LUM, were all technically verified as biomarker candidates. These proteins are likely truly differentially expressed in our cohort and need to be validated in external cohorts to assess their generalisability. However, verifying their differential expression does not necessarily make them good biomarkers. The effect sizes for the verified proteins were generally small and would probably not be clinically useful for diagnostic purposes. However, these proteins are probably involved in PD related mechanisms that could be further mechanistically explored and could result in novel drug targets. If other candidate biomarkers from the discovery study are validated with other methods, for example with immunoassays, they would potentially be useful as biomarkers. The highest accuracy in the cross validated machine learning analysis was found for F2R Like Trypsin Receptor 1 (F2RL1) at 76%. Ideally, for clinical diagnostic utility the accuracy needs to be higher than 76%. It was however demonstrated in this thesis that with 10 proteins, an accuracy >90% could be achieved, which is on par with clinical diagnostic accuracy [219], and therefore perhaps a panel of several proteins would be useful as a diagnostic tool.

Sex differences in PD have been described in the literature, particularly in terms of incidence and clinical presentation, although biological differences are much less studied [159]. Sex-specific analyses of the proteomic data generated interesting results, where the top biomarker candidates were quite different for male and female PD patients. This could be partly attributed to statistics, such as a high false discovery rate which is commonly seen in large scale proteomic studies. However, looking at the p-value histogram for female PD diagnostic biomarkers, one would expect approximately half of the statistically significant results to be true, yet few candidate biomarkers overlapped between male and female PD patients. Moreover, a greater number of significantly changed proteins were seen in female PD patients with regards to diagnostic biomarkers, whereas more candidate biomarkers were found for male patients that were associated with cognitive decline. It was hypothesised that part of this difference was due to cohort heterogeneity, however after a closer look at several clinical and demographic variables, only smaller differences were found between the male and female cohort. Likely some of these differences are due to biological differences. There is sparse literature on biological differences between male and female PD, however there are some studies that for example found an association between early menopause and increased risk of developing PD in females [158], which suggests there is a hormonal component to the pathology.

One of the most well studied neurodegenerative biomarkers today is neurofilament light chain (NfL). Plasma NfL levels were quantified in our cohort, in order to compare our proteomic findings with a

frequently bench marked plasma protein in PD, and to get an idea of how our cohort samples compare with the literature. NfL was indeed elevated in PD, and the increase was pronounced in the cognitively impaired patients. The effect size was attenuated after correcting for age, and only showed a modest elevation in PD and PDCI plasma. Moreover, NfL correlated with motor disease severity, even after age correction. The findings are in line with the current literature [220-222]. From our verified candidate proteins, APOA4 and LUM reflected conversion to PDD better than NfL, and the SAA1 peptide FFGHGAEDSLADQAANEWGR correlated better with MoCA scores than NfL. This suggests there most likely exists better biomarkers for PD diagnosis and cognition than NfL.

4.5 Assessment of the method

The protocol for the discovery study was heavily based on the protocol used by Ashton et al. [117] with a few minor modifications. The fractionation method was changed from OFFGEL strips separating samples by isoelectric point to spin columns where the bound sample was eluted using increasing concentrations of acetonitrile to fractionate the samples by hydrophobicity. This proved to be a less labour-intensive method that also yielded higher number of protein IDs and less technical variation compared with the OFFGEL method. Moreover, the Thermo Scientific Orbitrap Fusion Lumos instrument was used for the analysis, which in theory has improved sensitivity compared to the LTQ Orbitrap Velos used by Ashton et al. [117].

Over 20,000 protein IDs were identified in the discovery study at a 5% FDR, which is significantly more than the 500 proteins that were identified at a 1% FDR. Both approaches have their advantages and disadvantages. The convention is often to use 1% FDR [223, 224] in order to maximise true positive hits, while still detecting an adequate number of proteins. Although this approach gives higher confidence in the targets, one may miss out on many relevant targets by excluding false negatives. Plasma is a particularly complex sample matrix where protein abundances span a huge dynamic range. Plasma is considered to span at least 12 orders of magnitude, and the 12 most abundant proteins comprise 95% of the total protein abundance [225]. Therefore, plasma analysis often uses high abundant protein depletion steps, and prefractionation with gels or spin columns to maximise resolution. Albumin makes up more than half of the plasma proteins, followed by immunoglobulins, complement factors, iron binding proteins, apolipoproteins, and coagulation related proteins [225]. Comparatively, CNS derived proteins are of much lower concentration in plasma. The concentration of NfL for example is 10 orders of magnitude lower concentrations than albumin in plasma [84]. This is compared with for example CSF where the difference is 5 orders of magnitude [226, 227]. If the aim of the study is to generate biomarker candidates which need to be further verified and validated, it is preferential to generate a larger number of less robust candidates at the cost of increasing the number

of false positives. This strategy would reduce the risk of eliminating many potentially true positive hits, particularly among lower abundant proteins which more commonly have less confident peptide spectrum matches. Therefore a 5% FDR approach was selected, with a subsequent verification stage. However, a threshold was set where the protein needed to be quantified in at least 50% of the samples to be used in the downstream analysis, to mitigate the risk of too many false positives. This eliminated approximately 90% of all the hits.

When building our biomarker pipeline, a lack of verification experiments following proteomic biomarker candidates was noticed in the literature. There is a significant technical discrepancy between a proteomic approach where mass spectra from peptides in a fractionated sample are used as quantification, compared with immunoassays that utilise antibodies targeted to specific epitopes in native diluted sample biofluid. Parallel reaction monitoring (PRM) was used to verify the most promising biomarker candidate from the discovery study [138]. The results were relatively consistent with the discovery data for about half the proteins. A pronounced signal intensity variation was noted across samples, which was corrected for with total ion count and the QC samples that were run every 10 samples. The QC correction in this study was performed with external controls, and not internal controls, in order to keep the protocol more simple, and the samples as neat as possible. Ideally, to correct for the signal variation observed one would spike the samples with stable isotope standards to have an internal reference peptides to correct against [137, 228]. The PRM method used was effective for quantification of reproducible quality peptides. This meant that many of the less stringently identified biomarker candidates from the discovery study were not possible to verify with PRM, and other methods, such as immunoassays, need to be considered to verify the remaining candidates.

4.6 Comparison with CSF discovery studies.

What approach is taken for biomarker discovery is often dependent on the pathological hypothesis and clinical application. For example, if it believed a disease is isolated to one organ system, tissue or local biofluid, those tissues/fluids would be appropriate for biomarker discovery, as pathologically altered molecules would have a higher relative abundance near the source. There is a case to be made for discovery of neurological biomarkers in CSF, to then be replicated in plasma to improve its clinical utility. This may be a reasonable approach if a disease is believed to originate in the CNS, and some disease associated molecules leak or are transported into plasma from the CSF. For example, s100b, a common marker for TBI, is 500-fold higher in CSF than plasma following TBI [229]. It is an intracellular glial protein believed to be elevated following structural brain damage and is leaked into the plasma. Similarly, plasma proteins that directly reflect amyloid plaques and tau tangles in AD, such as p-Tau-

181, would be better to first establish in CSF, and later validate in plasma [230]. The aetiology of PD is poorly understood, and despite the core pathological hallmarks are in the brain, evidence suggests a key role of the periphery in early PD. Plasma was selected as the medium of choice in this study, not to miss any relevant peripheral proteins.

Nonetheless, to best identify candidate CNS derived PD biomarker candidates, large-scale studies on PD CSF should be performed. A few attempts have recently been made to both discover and validate CSF PD biomarkers. Naskar et al. [231] performed a label-free CSF proteomic study on PD, PDCI (PD with cognitive impairment) and control subjects, with approximately 10 participants per group. They identified 281 proteins, out of which 28 and 21 were differentially expressed in PDCI and PD, respectively, and managed to validate complement factor H in PDCI and Gelsolin in PD. Zhu et al. [137] performed an untargeted mass spectrometry experiment on PD, APD and controls, and identified peptides from 313 proteins. They validated their top hits with Multiple Reaction Monitoring (MRM) (n = approximately 30 per group) and found Leucine-rich α-2 glycoprotein 1 (LRG1) was elevated in both the profiling (discovery) study and the MRM study. Similar to our results, they too found elevated levels of complement factor C9 in PD patients in the discovery study, but noticed it varied greatly with sex. Marques et al. [232]performed an untargeted study using CSF from PD, MSA and controls (n = approximately 10 per group) and validated their findings using Selective Reaction Monitoring (SRM). However, the primary focus of their study was differentiating MSA from PD and controls. Rotunno et al. [233] used two cohorts of PD and controls (n = approximately 50 per group), and quantified CSF proteins using data independent acquisition (DIA) and validated 13 differentially expressed proteins in the two cohorts. Most recently Karayel et al. [234] quantified a total of over 1700 proteins from 2 cohorts PD and controls using a rectangular approach, and additionally included *LRRK2* carriers with and without PD (n = approximately 50 per group) using DIA. They validated osteomodulin and CD44 as differentially expressed in PD, as well as several HLA proteins in *LRRK2* PD patients.

An increasing number of groups are turning to data independent acquisition DIA for discovery studies. DIA allows for more complete MS2 spectra, and protein identification is done with a pre-constructed spectral library. This generally results in better scalability, and more accurate and reproducible results, and more accurate label free quantification [235]. Data dependent acquisition (DDA) sensitivity and quantification accuracy can however be improved with sample pre-fractionation and isobaric labelling respectively [236], although this results in more labour intensive studies. A general limitation of mass spectrometric discovery studies, is the poor quantification of lower abundant proteins in complex media. This is however a technical hurdle that might be overcome in the future with improved instruments, software, and spectral libraries [235, 237]. The most recent PD CSF study by Karayel et

al. [234] did for example not manage to quantify NfL or interleukins, which highlights the limitations of mass spectrometric techniques of today

Few biomarker candidates from the abovementioned CSF discovery studies, except for C9, overlapped with the plasma results in this thesis. This is possibly a reflection of most brain derived proteins that leak into plasma are low abundant, and difficult to quantify in with mass spectrometry. To some degree it could also reflect challenges in reproducing mass spectrometric data, and the stochastic nature of DDA generated data.

4.7 Limitations and future directions of discovery studies

One of the main objectives of this thesis was to identify plasma biomarkers that would reflect cognitive function and cognitive decline. One general limitation with the data analysis on disease severity and progression was that the plasma samples were not always taken from the time of diagnosis. The biomarker candidates associated with cognitive decline in Parkinson's Disease therefore do not necessarily reflect early changes that predict cognitive decline, but rather reflect what is seen in plasma in a patient that do decline more rapidly.

Not using baseline samples also introduced covariates such as dopaminergic medication and disease duration in the analyses. Given that disease duration, age, symptom severity, and medication were usually collinear, correcting for several of these would mean that interaction terms would need to be considered. This would generate overly complex linear models, and increase the risk of overfitting. Therefore, the data was generally only corrected for age and sex where appropriate, and covariates were often analysed separately to assess whether they had a significant effect on the biomarkers. The fact that most patients were on levodopa medication did mask some of the symptom severity, particularly UPDRS part III. Some studies assess the UPDRS score off medication, which results in more accurate motor assessments, but it is more difficult to implement in routine clinical practise [238]. Similar issues are true for non-levodopa medications. Patients with NMS, particularly depression and anxiety, are often on treatment which would mask some of the NMS severity. Moreover, many elderly individuals have several comorbidities, such as hypertension, cardiovascular disease, cancers, and diabetes, which both on their own and via their associated medications, likely affect the plasma proteome of the study individuals. It was attempted to select PD patients and controls with as few medications and comorbidities as possible, but it is often difficult to achieve owing to age of the study population and limited sample availability. Although baseline samples are ideal to analyse disease progression, they are less useful for cross sectional analyses. Many PD patients experience an exacerbation of symptom severity over time, and samples from a few years after diagnosis are generally better for assessing symptom severity and symptom phenotypes.

Another limitation of the study was a lack of an independent validation cohort. Instead, a large discovery cohort was used to increase the confidence of the biomarker candidates, particularly the low abundant ones. The proteins were then validated *in silico* through cross validation machine learning models. A subset of data was trained on itself, and then tested on another subset of data to assess its biomarker performance on new data. Just like a validation cohort, this tests the results on a new data, but does not account for other confounders such as cohort variation and technical variations. Some studies are opting for a "rectangular" biomarker discovery and validation approach which has some statistical benefits [135, 239]. With this approach a discovery study is performed in two separate cohorts, either with two real cohorts or with two artificially *in silico* constructed cohorts from one larger cohort. Overlapping significantly differentially expressed proteins are then considered validated biomarkers. This is an effective method as two large cohorts each generate high confidence biomarker candidates that are instantly validated in the same experiment. The trade-off is higher cost, more labour intensive setup, and the need of large cohort [135].

4.8 Complement study

Proteins related to the complement system have emerged as candidate plasma biomarkers in several proteomic studies [114] and the complement and coagulation cascade was the most implicated pathway in PD in the discovery study. As a validation study, proteins from the complement system were quantified in plasma from PD, HC, and 4R-Tauopathies (PSP or CBS). It was found that individuals with 4R-Tauopathies (PSP or CBS) had lower circulating C1q and C3 in plasma compared to both HC and PD. To our knowledge, these are novel findings and suggest plasma complement biomarkers may be able to differentiate APD from PD. Very few other targeted studies have looked at plasma complement factors in APD. Yamada et al. measured C4d and circulating immune complexes (CIC) to C1q in both plasma and CSF in PSP and PD, and found that C4d was increased in PSP plasma [240], and did not observe any changes in CIC to C1q in PSP. However, their cohort was quite small (6 PSP vs 14 PD), and they measured CIC bound C1q and not total C1q. Wang et al. measured CSF C3 and FH in PD, AD and Multiple System Atrophy (MSA) using a Luminex assay, and found that the C3 to FH ratio was decreased in individuals with MSA and could discriminate MSA from PD, AD and HC with high sensitivity and specificity [241]. They did not observe any changes in CSF C3 and FH between PD and HC. Other targeted biomarker studies in PSP and CBS have focused on core pathological changes. Being tauopathies, both total Tau and phosphorylated-Tau have been measured in 4R-Tauopathies CSF, but without consistent findings [242, 243]. Why peripheral Tau levels are altered in other Tau-related disorders like Alzheimer's disease (AD) but not in 4R-Tauopathies is unclear and hypothesised to be due to different protein isoforms or the involvement of beta-amyloid. Neurofilament light has probably been the most successful biomarker in differentiating PD from APD, both in plasma and CSF,

although it is less accurate in differentiating between different APDs [92]. For differentiation between APDs, perhaps a panel of proteins with C3 or C1q included may be useful. Magdanilou et al. showed that a panel of 9 different neurodegenerative markers in CSF could differentiate not only PD from APDs, but also different APDs from each other[244]. It is particularly interesting C1q was most significantly changed in the 4R-Tauopathies in this study, since both Tau and C1q have been considered two of the major contributors in AD neurodegeneration in some studies, and immunotherapies with bispecific antibodies against Tau and C1q are currently being developed[245].

Interestingly, no differences in any of the measured complement factors in plasma were observed between PD and HC using immunoassays. This was surprising, given the number of untargeted proteomic studies in PD plasma or serum finding changes in complement proteins. These studies report for example changes in levels of C5a[246], C3[53, 247, 248], C5, C4b and C1 inhibitor[52]. Furthermore, Alberio et al. performed an automated proteomic literature analysis on PD plasma with a subsequent experimental validation and concluded that both C3 and FH were viable PD plasma biomarkers [248], and Goldknopf et al. found nine protein spots related to the complement system on 2D-gels differentiating PD from other neurodegenerative diseases [247]. However, a major caveat is again the technical gap between proteomic discovery studies and immunoassay validation methods. As was seen in the PRM study, only one of the C9 peptides was elevated in PD plasma, which would for example not be detected with a Luminex assay if the antibody epitope binds to a different part of the C9 protein.

Only a few studies have measured complement factors in PD plasma in a targeted manner. Sun et al. used a scatter immune turbidity method to measure serum C3 and C4 in PD and HC [249], where they found no group differences. However, they noticed that female PD patients had higher C3 and C4 than male PD patients and female HC, and that male PD patients had lower C3 and C4 compared to male HC, and moreover, they saw both C3 and C4 were decreased with certain NMS. In our study, we did not find these sex-specific differences on a group level, increases in C3 with NMS severity for female PD patients were observed. Veselý et al. performed a longitudinal study in PD patients, where higher levels of serum C3 and C4 measured with nephelometry led to worsening in NMS and quality of life at a 2 year follow up [115]. Several of these patients also had increased levels of C3 and C4 at the follow up. Dufek et al. measured C1q, C1 inhibitor, C3, C4 and MBL in PD serum, and noticed lower MBL in a few of the PD patients [250]. Most of the proteins quantified in our study, have not been quantified in PD (or APD) plasma previously using immunoassays, including C4b, C5, C5a, C9, FB, FD, FH, and FI.

One inherent advantage of studying the complement system in blood is that all components are present, and activity can be studied *in vitro*. The 50% complement haemolytic assay (CH50) takes advantage of this. By adding serum to antibody-coated erythrocytes, the classical pathway is activated by the binding of C1q to the antibodies and ends with MAC pores forming on the erythrocytes causing haemolysis which is measured with absorbance. There was no difference in serum CH50 in PD versus controls. However, male patients higher levels of mental fatigue with higher classical complement activity, which suggests an inflammatory component might be involved in fatigue. CH50 assays have been performed for a few neuroinflammatory diseases, most notably for neuromyelitis optica, where lower CH50 was observed[251].

There were a few limitations to the complement study. Sera was not available for the BIOPARK study participants, and the CH50 assay was performed in a separate cohort (AETIONOMY), which meant complement levels and activity was not measured in the same individuals. However, the study still provides insight into how complement levels and activity are altered in the blood of individuals with PD, but not necessarily their relation. The 4R-Tauopathies group was small, heterogeneous, and could not be age and sex matched to the PD and HC groups. However, PSP and CBS are rare disorders where these complement proteins had not been measured previously, hence including an APD group still added value to the study. The diagnoses of PD, CBS and PSP were all probable, and not definitive as there was no *post-mortem* histopathological data. Future studies using pathologically confirmed APD cases could perhaps explain some of the heterogeneity of the 4R-Tauopathies group. Furthermore, the dynamic range for all complement proteins was quite large and overlapping between groups, which makes it difficult to set useful thresholds to use these proteins as biomarkers. Although the PD and HC groups were age matched, there were more male PD patients and female HC in our cohorts. However, this was corrected for in the regression models, and sex-specific analyses were also performed. Finally, we did not correct for multiple comparisons to avoid type 2 errors, as there were significant interactions between multiple complement proteins and clinical parameters respectively. However, further studies are needed to validate these findings.

To summarise, 4R-Tauopathies appear to have lower plasma C1q and C3 compared to PD and HC, and C3 correlates positively with several NMS in female PD patients. However, a group level difference in complement concentrations and complement activity was not found in PD patients.

4.9 Overall conclusion

In conclusion, the discovery study generated many novel plasma biomarker candidates for both PD diagnosis and cognition, and showed unbiased identification of new plasma biomarker candidates with mass spectrometry is indeed possible. The pipeline was successful in verifying several plasma proteins using PRM from the discovery study. Several novel biomarkers from the discovery study, and some from the verification study, showed superior biomarker performance to NfL. Generally, the pipeline approach worked well for robust verifiable peptides, and the discovery study did generate many new promising biomarker candidates for future studies to validate.

References

1. Deuschl, G., et al., *The burden of neurological diseases in Europe: an analysis for the Global Burden of Disease Study 2017.* The Lancet Public Health, 2020. **5**(10): p. e551-e567.
2. Postuma, R.B., et al., *MDS clinical diagnostic criteria for Parkinson's disease.* Mov Disord, 2015. **30**(12): p. 1591-601.
3. Geut, H., et al., *Neuropathological correlates of parkinsonian disorders in a large Dutch autopsy series.* Acta Neuropathologica Communications, 2020. **8**(1): p. 39.
4. Dickson, D.W., *Neuropathology of Parkinson disease.* Parkinsonism Relat Disord, 2018. **46 Suppl 1**(Suppl 1): p. S30-s33.
5. Chaudhuri, K.R., et al., *The burden of non-motor symptoms in Parkinson's disease using a self-completed non-motor questionnaire: a simple grading system.* Parkinsonism Relat Disord, 2015. **21**(3): p. 287-91.
6. Schapira, A.H.V., K.R. Chaudhuri, and P. Jenner, *Non-motor features of Parkinson disease.* Nat Rev Neurosci, 2017. **18**(7): p. 435-450.
7. Bloem, B.R., M.S. Okun, and C. Klein, *Parkinson's disease.* Lancet, 2021. **397**(10291): p. 2284-2303.
8. Berg, D., et al., *Prodromal Parkinson disease subtypes — key to understanding heterogeneity.* Nature Reviews Neurology, 2021. **17**(6): p. 349-361.
9. Luo, L., et al., *Motor phenotype classification in moderate to advanced PD in BioFIND study.* Parkinsonism Relat Disord, 2019. **65**: p. 178-183.
10. Nutt, J.G., *Motor subtype in Parkinson's disease: Different disorders or different stages of disease?* Movement Disorders, 2016. **31**(7): p. 957-961.
11. Fereshtehnejad, S.M., et al., *New Clinical Subtypes of Parkinson Disease and Their Longitudinal Progression: A Prospective Cohort Comparison With Other Phenotypes.* JAMA Neurol, 2015. **72**(8): p. 863-73.
12. Bohnen, N.I. and R.B. Postuma, *Body-first versus brain-first biological subtyping of Parkinson's disease.* Brain, 2020. **143**(10): p. 2871-2873.
13. Tysnes, O.-B. and A. Storstein, *Epidemiology of Parkinson's disease.* Journal of Neural Transmission, 2017. **124**(8): p. 901-905.
14. Pringsheim, T., et al., *The prevalence of Parkinson's disease: A systematic review and meta-analysis.* Movement Disorders, 2014. **29**(13): p. 1583-1590.
15. Cerri, S., L. Mus, and F. Blandini, *Parkinson's Disease in Women and Men: What's the Difference?* Journal of Parkinson's Disease, 2019. **9**: p. 501-515.

16. Noyce, A.J., et al., *Meta-analysis of early nonmotor features and risk factors for Parkinson disease.* Ann Neurol, 2012. **72**(6): p. 893-901.
17. Verstraeten, A., J. Theuns, and C. Van Broeckhoven, *Progress in unraveling the genetic etiology of Parkinson disease in a genomic era.* Trends in Genetics, 2015. **31**(3): p. 140-149.
18. Nalls, M.A., et al., *Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies.* The Lancet Neurology, 2019. **18**(12): p. 1091-1102.
19. Nalls, M.A., et al., *Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease.* Nature Genetics, 2014. **46**(9): p. 989-993.
20. Blauwendraat, C., M.A. Nalls, and A.B. Singleton, *The genetic architecture of Parkinson's disease.* The Lancet Neurology, 2020. **19**(2): p. 170-178.
21. Trinh, J., et al., *Genotype-phenotype relations for the Parkinson's disease genes SNCA, LRRK2, VPS35: MDSGene systematic review.* Mov Disord, 2018. **33**(12): p. 1857-1870.
22. Cherian, A. and K.P. Divya, *Genetics of Parkinson's disease.* Acta Neurologica Belgica, 2020. **120**(6): p. 1297-1305.
23. Tolosa, E., et al., *LRRK2 in Parkinson disease: challenges of clinical trials.* Nat Rev Neurol, 2020. **16**(2): p. 97-107.
24. Ge, P., V.L. Dawson, and T.M. Dawson, *PINK1 and Parkin mitochondrial quality control: a source of regional vulnerability in Parkinson's disease.* Molecular Neurodegeneration, 2020. **15**(1): p. 20.
25. Ovallath, S. and B. Sulthana, *Levodopa: History and Therapeutic Applications.* Annals of Indian Academy of Neurology, 2017. **20**(3): p. 185-189.
26. Armstrong, M.J. and M.S. Okun, *Diagnosis and Treatment of Parkinson Disease: A Review.* JAMA, 2020. **323**(6): p. 548-560.
27. Seppi, K., et al., *Update on treatments for nonmotor symptoms of Parkinson's disease-an evidence-based medicine review.* Mov Disord, 2019. **34**(2): p. 180-198.
28. Stoker, T.B. and R.A. Barker, *Recent developments in the treatment of Parkinson's Disease.* F1000Res, 2020. **9**.
29. Ntetsika, T., P.E. Papathoma, and I. Markaki, *Novel targeted therapies for Parkinson's disease.* Mol Med, 2021. **27**(1): p. 17.
30. Kalia, L.V. and A.E. Lang, *Parkinson's disease.* Lancet, 2015. **386**(9996): p. 896-912.
31. Raza, C., R. Anjum, and N.u.A. Shakeel, *Parkinson's disease: Mechanisms, translational models and management strategies.* Life Sciences, 2019. **226**: p. 77-90.

32. Braak, H., et al., *Staging of brain pathology related to sporadic Parkinson's disease.* Neurobiology of Aging, 2003. **24**(2): p. 197-211.
33. Doty, R.L., *Olfactory dysfunction in Parkinson disease.* Nature Reviews Neurology, 2012. **8**(6): p. 329-339.
34. Henderson, M.X., J.Q. Trojanowski, and V.M. Lee, *α-Synuclein pathology in Parkinson's disease and related α-synucleinopathies.* Neurosci Lett, 2019. **709**: p. 134316.
35. Kuzkina, A., et al., *Diagnostic value of skin RT-QuIC in Parkinson's disease: a two-laboratory study.* npj Parkinson's Disease, 2021. **7**(1): p. 99.
36. Kim, S., et al., *Transneuronal Propagation of Pathologic α-Synuclein from the Gut to the Brain Models Parkinson's Disease.* Neuron, 2019. **103**(4): p. 627-641.e7.
37. Burré, J., M. Sharma, and T.C. Südhof, *Cell Biology and Pathophysiology of α-Synuclein.* Cold Spring Harb Perspect Med, 2018. **8**(3).
38. Zheng, H., et al., *α-Synuclein in Parkinson's Disease: Does a Prion-Like Mechanism of Propagation from Periphery to the Brain Play a Role?* Neuroscientist, 2021. **27**(4): p. 367-387.
39. *Lewy body pathology is seen in cutaneous nerves and can be detected in skin biopsy samples.* Nature Clinical Practice Neurology, 2008. **4**(12): p. 640-640.
40. Liddle, R.A., *Parkinson's disease from the gut.* Brain Res, 2018. **1693**(Pt B): p. 201-206.
41. Duda, J.E., *Olfactory system pathology as a model of Lewy neurodegenerative disease.* J Neurol Sci, 2010. **289**(1-2): p. 49-54.
42. Malpartida, A.B., et al., *Mitochondrial Dysfunction and Mitophagy in Parkinson’s Disease: From Mechanism to Therapy.* Trends in Biochemical Sciences, 2021. **46**(4): p. 329-343.
43. Kim, S., et al., *Dysregulation of mitochondria-lysosome contacts by GBA1 dysfunction in dopaminergic neuronal models of Parkinson's disease.* Nat Commun, 2021. **12**(1): p. 1807.
44. Tansey, M.G., et al., *Inflammation and immune dysfunction in Parkinson disease.* Nature Reviews Immunology, 2022.
45. Tan, E.-K., et al., *Parkinson disease and the immune system — associations, mechanisms and therapeutics.* Nature Reviews Neurology, 2020. **16**(6): p. 303-318.
46. Li, X., J. Sundquist, and K. Sundquist, *Subsequent risks of Parkinson disease in patients with autoimmune and related disorders: a nationwide epidemiological study from Sweden.* Neurodegener Dis, 2012. **10**(1-4): p. 277-84.
47. Gerhard, A., et al., *In vivo imaging of microglial activation with [11C](R)-PK11195 PET in idiopathic Parkinson's disease.* Neurobiol Dis, 2006. **21**(2): p. 404-12.

48. McGeer, P.L., et al., *Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains.* Neurology, 1988. **38**(8): p. 1285-91.
49. Baird, J.K., et al., *The key role of T cells in Parkinson's disease pathogenesis and therapy.* Parkinsonism Relat Disord, 2019. **60**: p. 25-31.
50. Kitamura, Y., et al., *Proteomic Profiling of Exosomal Proteins for Blood-based Biomarkers in Parkinson's Disease.* Neuroscience, 2018. **392**: p. 121-128.
51. Zhao, X., et al., *Proteome analysis of the sera from Chinese Parkinson's disease patients.* Neurosci Lett, 2010. **479**(2): p. 175-9.
52. Zhang, X., et al., *Quantitative proteomic analysis of serum proteins in patients with Parkinson's disease using an isobaric tag for relative and absolute quantification labeling, two-dimensional liquid chromatography, and tandem mass spectrometry.* Analyst, 2012. **137**(2): p. 490-5.
53. Chiu, C.C., et al., *Increased Rab35 expression is a potential biomarker and implicated in the pathogenesis of Parkinson's disease.* Oncotarget, 2016. **7**(34): p. 54215-54227.
54. Carpanini, S.M., M. Torvell, and B.P. Morgan, *Therapeutic Inhibition of the Complement System in Diseases of the Central Nervous System.* Frontiers in Immunology, 2019. **10**.
55. Singh, A., L. Zhi, and H. Zhang, *LRRK2 and mitochondria: Recent advances and current views.* Brain Res, 2019. **1702**: p. 96-104.
56. Rocha, E.M., B. De Miranda, and L.H. Sanders, *Alpha-synuclein: Pathology, mitochondrial dysfunction and neuroinflammation in Parkinson's disease.* Neurobiol Dis, 2018. **109**(Pt B): p. 249-257.
57. Navarro-Romero, A., M. Montpeyó, and M. Martinez-Vicente, *The Emerging Role of the Lysosome in Parkinson's Disease.* Cells, 2020. **9**(11).
58. Alcalay, R.N., et al., *Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations.* Brain, 2015. **138**(Pt 9): p. 2648-58.
59. Fasano, A., et al., *Gastrointestinal dysfunction in Parkinson's disease.* The Lancet Neurology, 2015. **14**(6): p. 625-639.
60. Svensson, E., et al., *Vagotomy and subsequent risk of Parkinson's disease.* Ann Neurol, 2015. **78**(4): p. 522-9.
61. Aarsland, D., et al., *Cognitive decline in Parkinson disease.* Nat Rev Neurol, 2017. **13**(4): p. 217-231.
62. Roheger, M., E. Kalbe, and I. Liepelt-Scarfone, *Progression of Cognitive Decline in Parkinson's Disease.* J Parkinsons Dis, 2018. **8**(2): p. 183-193.

63. Aarsland, D., J. Zaccai, and C. Brayne, *A systematic review of prevalence studies of dementia in Parkinson's disease.* Mov Disord, 2005. **20**(10): p. 1255-63.
64. Aarsland, D., et al., *Parkinson disease-associated cognitive impairment.* Nature Reviews Disease Primers, 2021. **7**(1): p. 47.
65. Poewe, W., et al., *Diagnosis and management of Parkinson's disease dementia.* International journal of clinical practice, 2008. **62**(10): p. 1581-1587.
66. Walker, Z., et al., *Lewy body dementias.* Lancet (London, England), 2015. **386**(10004): p. 1683-1697.
67. Jellinger, K.A. and A.D. Korczyn, *Are dementia with Lewy bodies and Parkinson's disease dementia the same disease?* BMC Medicine, 2018. **16**(1): p. 34.
68. Gonzalez, M.C., et al., *Association of Plasma p-tau181 and p-tau231 Concentrations With Cognitive Decline in Patients With Probable Dementia With Lewy Bodies.* JAMA Neurology, 2022. **79**(1): p. 32-37.
69. Pilotto, A., et al., *Plasma Neurofilament Light Chain Predicts Cognitive Progression in Prodromal and Clinical Dementia with Lewy Bodies.* J Alzheimers Dis, 2021. **82**(3): p. 913-919.
70. Bargar, C., et al., *Streamlined alpha-synuclein RT-QuIC assay for various biospecimens in Parkinson's disease and dementia with Lewy bodies.* Acta Neuropathologica Communications, 2021. **9**(1): p. 62.
71. Delenclos, M., et al., *Biomarkers in Parkinson's disease: Advances and strategies.* Parkinsonism & related disorders, 2016. **22 Suppl 1**(Suppl 1): p. S106-S110.
72. Mahlknecht, P., et al., *Prodromal Parkinson's disease: hype or hope for disease-modification trials?* Translational Neurodegeneration, 2022. **11**(1): p. 11.
73. Postuma, R.B., et al., *Parkinson risk in idiopathic REM sleep behavior disorder: preparing for neuroprotective trials.* Neurology, 2015. **84**(11): p. 1104-13.
74. Jennings, D., et al., *Conversion to Parkinson Disease in the PARS Hyposmic and Dopamine Transporter-Deficit Prodromal Cohort.* JAMA Neurol, 2017. **74**(8): p. 933-940.
75. Ravina, B., et al., *Dopamine transporter imaging is associated with long-term outcomes in Parkinson's disease.* Movement disorders : official journal of the Movement Disorder Society, 2012. **27**(11): p. 1392-1397.
76. Summerfield, C., et al., *Structural brain changes in Parkinson disease with dementia: a voxel-based morphometry study.* Archives of neurology, 2005. **62**(2): p. 281-285.
77. Ryman, S.G. and K.L. Poston, *MRI biomarkers of motor and non-motor symptoms in Parkinson's disease.* Parkinsonism Relat Disord, 2020. **73**: p. 85-93.

78. Nabais, M.F., et al., *Meta-analysis of genome-wide DNA methylation identifies shared associations across neurodegenerative disorders.* Genome Biol, 2021. **22**(1): p. 90.
79. Parnetti, L., et al., *CSF and blood biomarkers for Parkinson's disease.* The Lancet. Neurology, 2019. **18**(6): p. 573-586.
80. Stöppler, M.C. *Medical Definition of Biofluid*. 2021 3/29/2021 [cited 2021; Available from: https://www.medicinenet.com/biofluid/definition.htm#:~:text=Biofluid%3A%20A%20biological%20fluid.,as%20blister%20or%20cyst%20fluid).
81. Zamir, M., et al., *Biofluid mechanics of special organs and the issue of system control. Sixth International Bio-Fluid Mechanics Symposium and Workshop, March 28-30, 2008 Pasadena, California.* Ann Biomed Eng, 2010. **38**(3): p. 1204-15.
82. Chaudhry, R., J.H. Miao, and A. Rehman, *Physiology, Cardiovascular*, in *StatPearls*. 2022, StatPearls Publishing

83. Ozdowski, L. and V. Gupta, *Physiology, Lymphatic System*, in *StatPearls*. 2022, StatPearls Publishing

84. Ashton, N.J., et al., *A multicentre validation study of the diagnostic value of plasma neurofilament light.* Nature Communications, 2021. **12**(1): p. 3400.
85. Thijssen, E.H., et al., *Plasma phosphorylated tau 217 and phosphorylated tau 181 as biomarkers in Alzheimer's disease and frontotemporal lobar degeneration: a retrospective diagnostic performance study.* Lancet Neurol, 2021. **20**(9): p. 739-752.
86. Li, Y., et al., *Validation of Plasma Amyloid-β 42/40 for Detecting Alzheimer Disease Amyloid Plaques.* Neurology, 2022. **98**(7): p. e688-e699.
87. Teunissen, C.E., et al., *Blood-based biomarkers for Alzheimer's disease: towards clinical implementation.* Lancet Neurol, 2022. **21**(1): p. 66-77.
88. Blennow, K. and H. Zetterberg, *Biomarkers for Alzheimer's disease: current status and prospects for the future.* J Intern Med, 2018. **284**(6): p. 643-663.
89. Mielke, M.M., et al., *Comparison of Plasma Phosphorylated Tau Species With Amyloid and Tau Positron Emission Tomography, Neurodegeneration, Vascular Pathology, and Cognitive Outcomes.* JAMA Neurol, 2021. **78**(9): p. 1108-1117.
90. Mavroudis, I.A., et al., *A meta-analysis on CSF neurogranin levels for the diagnosis of Alzheimer's disease and mild cognitive impairment.* Aging Clin Exp Res, 2020. **32**(9): p. 1639-1646.

91. Ashton, N.J., et al., *An update on blood-based biomarkers for non-Alzheimer neurodegenerative disorders.* Nat Rev Neurol, 2020. **16**(5): p. 265-284.
92. Gaetani, L., et al., *Neurofilament light chain as a biomarker in neurological disorders.* J Neurol Neurosurg Psychiatry, 2019. **90**(8): p. 870-881.
93. Bäckström, D., et al., *NfL as a biomarker for neurodegeneration and survival in Parkinson disease.* Neurology, 2020. **95**(7): p. e827-e838.
94. Parnetti, L., et al., *CSF and blood biomarkers for Parkinson's disease.* The Lancet Neurology, 2019. **18**(6): p. 573-586.
95. Eusebi, P., et al., *Diagnostic utility of cerebrospinal fluid α-synuclein in Parkinson's disease: A systematic review and meta-analysis.* Movement Disorders, 2017. **32**(10): p. 1389-1400.
96. Sako, W., et al., *Reduced alpha-synuclein in cerebrospinal fluid in synucleinopathies: Evidence from a meta-analysis.* Movement Disorders, 2014. **29**(13): p. 1599-1605.
97. Gao, L., et al., *Cerebrospinal fluid alpha-synuclein as a biomarker for Parkinson's disease diagnosis: a systematic review and meta-analysis.* Int J Neurosci, 2015. **125**(9): p. 645-54.
98. Zhou, B., et al., *The Diagnostic and Differential Diagnosis Utility of Cerebrospinal Fluid α-Synuclein Levels in Parkinson's Disease: A Meta-Analysis.* Parkinson’s Disease, 2015. **2015**: p. 567386.
99. Parnetti, L., et al., *Differential role of CSF alpha-synuclein species, tau, and Aβ42 in Parkinson's Disease.* Frontiers in Aging Neuroscience, 2014. **6**.
100. Arawaka, S., et al., *Mechanisms underlying extensive Ser129-phosphorylation in α-synuclein aggregates.* Acta Neuropathol Commun, 2017. **5**(1): p. 48.
101. Fairfoul, G., et al., *Alpha-synuclein RT-QuIC in the CSF of patients with alpha-synucleinopathies.* Ann Clin Transl Neurol, 2016. **3**(10): p. 812-818.
102. Candelise, N., et al., *Towards an improved early diagnosis of neurodegenerative diseases: the emerging role of in vitro conversion assays for protein amyloids.* Acta Neuropathol Commun, 2020. **8**(1): p. 117.
103. Rossi, M., et al., *Ultrasensitive RT-QuIC assay with high sensitivity and specificity for Lewy body-associated synucleinopathies.* Acta Neuropathol, 2020. **140**(1): p. 49-62.
104. Quadalti, C., et al., *Neurofilament light chain and α-synuclein RT-QuIC as differential diagnostic biomarkers in parkinsonisms and related syndromes.* npj Parkinson's Disease, 2021. **7**(1): p. 93.
105. Kwon, E.H., et al., *Update on CSF Biomarkers in Parkinson’s Disease.* Biomolecules, 2022. **12**(2): p. 329.

106. Lerche, S., et al., *CSF NFL in a Longitudinally Assessed PD Cohort: Age Effects and Cognitive Trajectories.* Movement Disorders, 2020. **35**(7): p. 1138-1144.
107. Aamodt, W.W., et al., *Neurofilament Light Chain as a Biomarker for Cognitive Decline in Parkinson Disease.* Mov Disord, 2021. **36**(12): p. 2945-2950.
108. Bougea, A., et al., *Plasma alpha-synuclein levels in patients with Parkinson's disease: a systematic review and meta-analysis.* Neurological Sciences, 2019. **40**(5): p. 929-938.
109. Lin, C.-H., et al., *Plasma Biomarkers Differentiate Parkinson's Disease From Atypical Parkinsonism Syndromes.* Frontiers in Aging Neuroscience, 2018. **10**.
110. Qin, X.-Y., et al., *Aberrations in Peripheral Inflammatory Cytokine Levels in Parkinson Disease: A Systematic Review and Meta-analysis.* JAMA Neurology, 2016. **73**(11): p. 1316-1324.
111. Weisskopf, M.G., et al., *Plasma urate and risk of Parkinson's disease.* Am J Epidemiol, 2007. **166**(5): p. 561-7.
112. Cortese, M., et al., *Urate and the risk of Parkinson's disease in men and women.* Parkinsonism Relat Disord, 2018. **52**: p. 76-82.
113. Schwarzschild, M.A., et al., *Effect of Urate-Elevating Inosine on Early Parkinson Disease Progression: The SURE-PD3 Randomized Clinical Trial.* Jama, 2021. **326**(10): p. 926-939.
114. Chelliah, S.S., et al., *Identification of blood-based biomarkers for diagnosis and prognosis of Parkinson's disease: A systematic review of proteomics studies.* Ageing Research Reviews, 2022. **73**: p. 101514.
115. Veselý, B., et al., *Interleukin 6 and complement serum level study in Parkinson's disease.* J Neural Transm (Vienna), 2018. **125**(5): p. 875-881.
116. Lewczuk, P., et al., *Cerebrospinal fluid and blood biomarkers for neurodegenerative dementias: An update of the Consensus of the Task Force on Biological Markers in Psychiatry of the World Federation of Societies of Biological Psychiatry.* World Journal of Biological Psychiatry, 2018. **19**(4): p. 244-328.
117. Ashton, N.J., et al., *A plasma protein classifier for predicting amyloid burden for preclinical Alzheimer's disease.* Sci Adv, 2019. **5**(2): p. eaau7220.
118. Williams-Gray, C.H., et al., *Evolution of cognitive dysfunction in an incident Parkinson's disease cohort.* Brain, 2007. **130**(7): p. 1787-1798.
119. Lin, C.-H. and R.-M. Wu, *Biomarkers of cognitive decline in Parkinson's disease.* Parkinsonism & Related Disorders, 2015. **21**(5): p. 431-443.
120. Ibarretxe-Bilbao, N., et al., *MRI and cognitive impairment in Parkinson's disease.* Movement Disorders, 2009. **24**(S2): p. S748-S753.

121. Yong, S., et al., *A comparison of cerebral glucose metabolism in Parkinson's disease, Parkinson's disease dementia and dementia with Lewy bodies.* European Journal of Neurology, 2007. **14**(12): p. 1357-1362.
122. Delgado-Alvarado, M., et al., *Biomarkers for dementia and mild cognitive impairment in Parkinson's disease.* Movement disorders : official journal of the Movement Disorder Society, 2016. **31**(6): p. 861-881.
123. Lim, X., et al., *The diagnostic utility of cerebrospinal fluid alpha-synuclein analysis in dementia with Lewy bodies–a systematic review and meta-analysis.* Parkinsonism & related disorders, 2013. **19**(10): p. 851-858.
124. Posavi, M., et al., *Characterization of Parkinson's disease using blood-based biomarkers: A multicohort proteomic analysis.* Plos Medicine, 2019. **16**(10).
125. Markaki, I., et al., *Cerebrospinal Fluid Levels of Kininogen-1 Indicate Early Cognitive Impairment in Parkinson's Disease.* Mov Disord, 2020. **35**(11): p. 2101-2106.
126. Jankovska, E., et al., *Affinity depletion versus relative protein enrichment: a side-by-side comparison of two major strategies for increasing human cerebrospinal fluid proteome coverage.* Clin Proteomics, 2019. **16**: p. 9.
127. Huang, Y., et al., *Serum NFL discriminates Parkinson disease from essential tremor and reflect motor and cognition severity.* BMC Neurology, 2022. **22**(1): p. 39.
128. Huang, J., et al., *Inflammation-related plasma and CSF biomarkers for multiple sclerosis.* Proceedings of the National Academy of Sciences, 2020. **117**(23): p. 12952.
129. Lynch, H.E., et al., *Development and implementation of a proficiency testing program for Luminex bead-based cytokine assays.* J Immunol Methods, 2014. **409**: p. 62-71.
130. Kiddle, S.J., et al., *Candidate blood proteome markers of Alzheimer's disease onset and progression: a systematic review and replication study.* J Alzheimers Dis, 2014. **38**(3): p. 515-31.
131. Bourmaud, A., S. Gallien, and B. Domon, *Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: Principle and applications.* Proteomics, 2016. **16**(15-16): p. 2146-59.
132. Yun, N., et al., *Anamorsin, a novel caspase-3 substrate in neurodegeneration.* J Biol Chem, 2014. **289**(32): p. 22183-95.
133. Nakayasu, E.S., et al., *Tutorial: best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation.* Nature Protocols, 2021. **16**(8): p. 3737-3760.

134. Tabb, D.L., et al., *Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry.* J Proteome Res, 2010. **9**(2): p. 761-76.
135. Geyer, P.E., et al., *Revisiting biomarker discovery by plasma proteomics.* Molecular systems biology, 2017. **13**(9): p. 942-942.
136. Mei, J., C. Desrosiers, and J. Frasnelli, *Machine Learning for the Diagnosis of Parkinson's Disease: A Review of Literature.* Frontiers in Aging Neuroscience, 2021. **13**.
137. Zhu, S., et al., *Alterations in Self-Aggregating Neuropeptides in Cerebrospinal Fluid of Patients with Parkinsonian Disorders.* J Parkinsons Dis, 2022. **12**(4): p. 1169-1189.
138. Rauniyar, N., *Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry.* Int J Mol Sci, 2015. **16**(12): p. 28566-81.
139. Zhang, Q., et al., *Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes.* J Exp Med, 2013. **210**(1): p. 191-203.
140. Carnielli, C.M., et al., *Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer.* Nature Communications, 2018. **9**(1): p. 3598.
141. Tofte, N., et al., *Early detection of diabetic kidney disease by urinary proteomics and subsequent intervention with spironolactone to delay progression (PRIORITY): a prospective observational study and embedded randomised placebo-controlled trial.* The Lancet Diabetes & Endocrinology, 2020. **8**(4): p. 301-312.
142. Fung, E.T., *A Recipe for Proteomics Diagnostic Test Development: The OVA1 Test, from Biomarker Discovery to FDA Clearance.* Clinical Chemistry, 2010. **56**(2): p. 327-329.
143. Lawton, M., et al., *Blood biomarkers with Parkinson's disease clusters and prognosis: The oxford discovery cohort.* Mov Disord, 2020. **35**(2): p. 279-287.
144. Goldman, J.G. and R. Postuma, *Premotor and nonmotor features of Parkinson's disease.* Curr Opin Neurol, 2014. **27**(4): p. 434-41.
145. Bäckström, D., et al., *Prediction and early biomarkers of cognitive decline in Parkinson disease and atypical parkinsonism: a population-based study.* Brain Communications, 2022. **4**(2).
146. Batzu, L., et al., *Plasma p-tau181, neurofilament light chain and association with cognition in Parkinson's disease.* npj Parkinson's Disease, 2022. **8**(1): p. 154.
147. Tanaka, T., et al., *Plasma proteomic biomarker signature of age predicts health and life span.* eLife, 2020. **9**: p. e61073.
148. Massey, S.C., et al., *Sex differences in health and disease: A review of biological sex differences relevant to cancer with a spotlight on glioma.* Cancer Lett, 2021. **498**: p. 178-187.

149. Rubin, J.B., *The spectrum of sex differences in cancer.* Trends in Cancer, 2022. **8**(4): p. 303-315.
150. Whitacre, C.C., *Sex differences in autoimmune disease.* Nature Immunology, 2001. **2**(9): p. 777-780.
151. Hannawa, K.K., J.L. Eliason, and G.R. Upchurch, Jr., *Gender differences in abdominal aortic aneurysms.* Vascular, 2009. **17 Suppl 1**(Suppl 1): p. S30-9.
152. Wickens, M.M., D.A. Bangasser, and L.A. Briand, *Sex Differences in Psychiatric Disease: A Focus on the Glutamate System.* Frontiers in Molecular Neuroscience, 2018. **11**.
153. Beam, C.R., et al., *Differences Between Women and Men in Incidence Rates of Dementia and Alzheimer's Disease.* J Alzheimers Dis, 2018. **64**(4): p. 1077-1083.
154. Mofrad, R.B., et al., *Sex differences in CSF biomarkers vary by Alzheimer disease stage and APOE ε4 genotype.* Neurology, 2020. **95**(17): p. e2378-e2388.
155. Tsiknia, A.A., et al., *Sex differences in plasma p-tau181 associations with Alzheimer's disease biomarkers, cognitive decline, and clinical progression.* Mol Psychiatry, 2022. **27**(10): p. 4314-4322.
156. Chiu, S.Y., et al., *Sex differences in dementia with Lewy bodies: Focused review of available evidence and future directions.* Parkinsonism & Related Disorders, 2023. **107**: p. 105285.
157. López-Cerdán, A., et al., *Unveiling sex-based differences in Parkinson's disease: a comprehensive meta-analysis of transcriptomic studies.* Biology of Sex Differences, 2022. **13**(1): p. 68.
158. Unda, S.R., et al., *State-of-the-art review of the clinical research on menopause and hormone replacement therapy association with Parkinson's disease: What meta-analysis studies cannot tell us.* Front Aging Neurosci, 2022. **14**: p. 971007.
159. Russillo, M.C., et al., *Sex Differences in Parkinson's Disease: From Bench to Bedside.* Brain Sci, 2022. **12**(7).
160. Nguyen, V.P., et al., *Sex Differences in the Level of Homocysteine in Alzheimer’s Disease and Parkinson’s Disease Patients: A Meta-Analysis.* Brain Sciences, 2023. **13**(1): p. 153.
161. Piscopo, P., et al., *A Sex Perspective in Neurodegenerative Diseases: microRNAs as Possible Peripheral Biomarkers.* Int J Mol Sci, 2021. **22**(9).
162. Janelidze, S., et al., *Plasma P-tau181 in Alzheimer's disease: relationship to other biomarkers, differential diagnosis, neuropathology and longitudinal progression to Alzheimer's dementia.* Nature Medicine, 2020. **26**.

163. Markaki, I., et al., *Euglycemia Indicates Favorable Motor Outcome in Parkinson's Disease.* Mov Disord, 2021. **36**(6): p. 1430-1434.
164. Goetz, C.G., et al., *Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results.* Mov Disord, 2008. **23**(15): p. 2129-70.
165. Nasreddine, Z.S., et al., *The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment.* J Am Geriatr Soc, 2005. **53**(4): p. 695-9.
166. Beck, A.T., et al., *Comparison of Beck Depression Inventories -IA and -II in psychiatric outpatients.* J Pers Assess, 1996. **67**(3): p. 588-97.
167. Montgomery, S.A. and M. Asberg, *A new depression scale designed to be sensitive to change.* Br J Psychiatry, 1979. **134**: p. 382-9.
168. Zigmond, A.S. and R.P. Snaith, *The hospital anxiety and depression scale.* Acta Psychiatr Scand, 1983. **67**(6): p. 361-70.
169. Johansson, B., et al., *A self-assessment questionnaire for mental fatigue and related symptoms after neurological disorders and injuries.* Brain Inj, 2010. **24**(1): p. 2-12.
170. Buysse, D.J., et al., *The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research.* Psychiatry Res, 1989. **28**(2): p. 193-213.
171. Jenkinson, C., et al., *The Parkinson's Disease Questionnaire (PDQ-39): development and validation of a Parkinson's disease summary index score.* Age Ageing, 1997. **26**(5): p. 353-7.
172. Chaudhuri, K.R., et al., *International multicenter pilot study of the first comprehensive self-completed nonmotor symptoms questionnaire for Parkinson's disease: the NMSQuest study.* Mov Disord, 2006. **21**(7): p. 916-23.
173. Tomlinson, C.L., et al., *Systematic review of levodopa dose equivalency reporting in Parkinson's disease.* Mov Disord, 2010. **25**(15): p. 2649-53.
174. Hentz, J.G., et al., *Simplified conversion method for unified Parkinson's disease rating scale motor examinations.* Movement disorders : official journal of the Movement Disorder Society, 2015. **30**(14): p. 1967-1970.
175. Dalrymple-Alford, J.C., et al., *The MoCA.* Well-suited screen for cognitive impairment in Parkinson disease, 2010. **75**(19): p. 1717-1725.
176. Borland, E., et al., *The Montreal Cognitive Assessment: Normative Data from a Large Swedish Population-Based Cohort.* J Alzheimers Dis, 2017. **59**(3): p. 893-901.
177. Zhang, X., et al., *Quantitative proteomic analysis of serum proteins in patients with Parkinson's disease using an isobaric tag for relative and absolute quantification labeling,*

two-dimensional liquid chromatography, and tandem mass spectrometry. The Analyst, 2012. **137**(2): p. 490-495.

178. Sun, C., et al., *Peripheral Humoral Immune Response Is Associated With the Non-motor Symptoms of Parkinson's Disease.* Frontiers in neuroscience, 2019. **13**: p. 1057-1057.
179. Höglinger, G.U., et al., *Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria.* Movement Disorders, 2017. **32**(6): p. 853-864.
180. Armstrong, M.J., et al., *Criteria for the diagnosis of corticobasal degeneration.* Neurology, 2013. **80**(5): p. 496-503.
181. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.
182. Duan, K.B., et al., *Multiple SVM-RFE for gene selection in cancer classification with expression data.* IEEE Trans Nanobioscience, 2005. **4**(3): p. 228-34.
183. Kanehisa, M., et al., *New approach for understanding genome variations in KEGG.* Nucleic Acids Research, 2018. **47**(D1): p. D590-D595.
184. Szklarczyk, D., et al., *The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets.* Nucleic Acids Research, 2020. **49**(D1): p. D605-D612.
185. Rissin, D.M., et al., *Single-molecule enzyme-linked immunosorbent assay detects serum proteins at subfemtomolar concentrations.* Nat Biotechnol, 2010. **28**(6): p. 595-9.
186. Ekdahl, K.N., et al., *Interpretation of Serological Complement Biomarkers in Disease.* Frontiers in Immunology, 2018. **9**(2237).
187. Lalezari, S., et al., *Correlation between endogenous VWF:Ag and PK parameters and bleeding frequency in severe haemophilia A subjects during three-times-weekly prophylaxis with rFVIII-FS.* Haemophilia, 2014. **20**(1): p. e15-e22.
188. Odden, M.C., et al., *Age and cystatin C in healthy adults: a collaborative study.* Nephrology Dialysis Transplantation, 2009. **25**(2): p. 463-469.
189. Hager, K., et al., *Fibrinogen and aging.* Aging (Milano), 1994. **6**(2): p. 133-8.
190. Folkersen, J., et al., *Circulating levels of pregnancy zone protein: Normal range and the influence of age and gender.* Clinica Chimica Acta, 1981. **110**(2): p. 139-145.
191. Yunice, A.A., et al., *Influence of age and sex on serum copper and ceruloplasmin levels.* J Gerontol, 1974. **29**(3): p. 277-81.
192. Rasmussen, K.L., *Plasma levels of apolipoprotein E, APOE genotype and risk of dementia and ischemic heart disease: A review.* Atherosclerosis, 2016. **255**: p. 145-155.

193. Green, H.F., S. Khosousi, and P. Svenningsson, *Plasma IL-6 and IL-17A Correlate with Severity of Motor and Non-Motor Symptoms in Parkinson's Disease.* J Parkinsons Dis, 2019. **9**(4): p. 705-709.

194. Oosterveld, L.P., et al., *CSF or serum neurofilament light added to α-Synuclein panel discriminates Parkinson's from controls.* Mov Disord, 2020. **35**(2): p. 288-295.

195. Mollenhauer, B., et al., *Validation of Serum Neurofilament Light Chain as a Biomarker of Parkinson's Disease Progression.* Mov Disord, 2020. **35**(11): p. 1999-2008.

196. Ma, L.Z., et al., *Serum Neurofilament Dynamics Predicts Cognitive Progression in de novo Parkinson's Disease.* J Parkinsons Dis, 2021. **11**(3): p. 1117-1127.

197. Zhu, Y., et al., *Association between plasma neurofilament light chain levels and cognitive function in patients with Parkinson's disease.* J Neuroimmunol, 2021. **358**: p. 577662.

198. Sampedro, F., et al., *Serum neurofilament light chain levels reflect cortical neurodegeneration in de novo Parkinson's disease.* Parkinsonism Relat Disord, 2020. **74**: p. 43-49.

199. Khalil, M., et al., *Serum neurofilament light levels in normal aging and their association with morphologic brain changes.* Nature Communications, 2020. **11**(1): p. 812.

200. Stamelou, M., N.P. Quinn, and K.P. Bhatia, *"Atypical" atypical parkinsonism: new genetic conditions presenting with features of progressive supranuclear palsy, corticobasal degeneration, or multiple system atrophy-a diagnostic guide.* Mov Disord, 2013. **28**(9): p. 1184-99.

201. Morgan, B.P., *Complement in the pathogenesis of Alzheimer's disease.* Semin Immunopathol, 2018. **40**(1): p. 113-124.

202. Barbour, R., et al., *Detection of ser 129 phosphorylated alpha-synuclein in human CSF using an ultra-sensitive immunoassay.* Neurodegenerative Diseases, 2017. **17**: p. 1540.

203. He, W.Q., et al., *Myosin light chain kinase is central to smooth muscle contraction and required for gastrointestinal motility in mice.* Gastroenterology, 2008. **135**(2): p. 610-20.

204. Chaudhuri, K.R. and A.H.V. Schapira, *Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment.* Lancet Neurology, 2009. **8**(5): p. 464-474.

205. Verma, A.K., et al., *Plasma Prolidase Activity and Oxidative Stress in Patients with Parkinson's Disease.* Parkinson’s Disease, 2015. **2015**: p. 598028.

206. Uhlén, M., et al., *Proteomics. Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.

207. Fernandez-Irigoyen, J., et al., *New insights into the human brain proteome: Protein expression profiling of deep brain stimulation target areas.* Journal of Proteomics, 2015. **127**: p. 395-405.
208. Licker, V., et al., *Proteomic analysis of human substantia nigra identifies novel candidates involved in Parkinson's disease pathogenesis.* Proteomics, 2014. **14**(6): p. 784-94.
209. Sack, G.H., *Serum amyloid A – a review.* Molecular Medicine, 2018. **24**(1): p. 46.
210. Hall, S., et al., *Inflammatory, degeneration and neuritic growth biomarkers predict cognitive decline and dementia in Parkinson's disease.* Alzheimer's & Dementia, 2021. **17**(S4): p. e053729.
211. Jang, S., et al., *Serum amyloid A1 is involved in amyloid plaque aggregation and memory decline in amyloid beta abundant condition.* Transgenic Res, 2019. **28**(5-6): p. 499-508.
212. Goodson, H.V. and E.M. Jonasson, *Microtubules and Microtubule-Associated Proteins.* Cold Spring Harb Perspect Biol, 2018. **10**(6).
213. Zelentsova, K., et al., *Protein S Regulates Neural Stem Cell Quiescence and Neurogenesis.* Stem Cells, 2017. **35**(3): p. 679-693.
214. Forero, D.A., et al., *Meta-analysis of Telomere Length in Alzheimer's Disease.* J Gerontol A Biol Sci Med Sci, 2016. **71**(8): p. 1069-73.
215. Lange, S., et al., *Obscurin determines the architecture of the longitudinal sarcoplasmic reticulum.* J Cell Sci, 2009. **122**(Pt 15): p. 2640-50.
216. Li, L., et al., *Relationship between Apolipoprotein Superfamily and Parkinson's Disease.* Chin Med J (Engl), 2017. **130**(21): p. 2616-2623.
217. Paslawski, W., et al., *α-synuclein–lipoprotein interactions and elevated ApoE level in cerebrospinal fluid from Parkinson's disease patients.* Proceedings of the National Academy of Sciences, 2019. **116**(30): p. 15226-15235.
218. IJsselstijn, L., et al., *Serum proteomics in amnestic mild cognitive impairment.* PROTEOMICS, 2013. **13**(16): p. 2526-2533.
219. Virameteekul, S., et al., *Clinical Diagnostic Accuracy of Parkinson's Disease: Where Do We Stand?* Movement Disorders. **n/a**(n/a).
220. Zhu, Y., et al., *Association between plasma neurofilament light chain levels and cognitive function in patients with Parkinson's disease.* Journal of Neuroimmunology, 2021. **358**: p. 577662.
221. Lin, Y.-S., et al., *Levels of plasma neurofilament light chain and cognitive function in patients with Alzheimer or Parkinson disease.* Scientific Reports, 2018. **8**(1): p. 17368.

222. Pilotto, A., et al., *Plasma NfL, clinical subtypes and motor progression in Parkinson's disease.* Parkinsonism Relat Disord, 2021. **87**: p. 41-47.
223. Burger, T., *Gentle Introduction to the Statistical Foundations of False Discovery Rate in Quantitative Proteomics.* J Proteome Res, 2018. **17**(1): p. 12-22.
224. Savitski, M.M., et al., *A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets.* Mol Cell Proteomics, 2015. **14**(9): p. 2394-404.
225. Hortin, G.L. and D. Sviridov, *The dynamic range problem in the analysis of the plasma proteome.* Journal of Proteomics, 2010. **73**(3): p. 629-636.
226. Murakami, K., et al., *Changes with aging of steroidal levels in the cerebrospinal fluid of women.* Maturitas, 1999. **33**(1): p. 71-80.
227. Bridel, C., et al., *Diagnostic Value of Cerebrospinal Fluid Neurofilament Light Protein in Neurology: A Systematic Review and Meta-analysis.* JAMA Neurology, 2019. **76**(9): p. 1035-1048.
228. Zhao, Y. and A.R. Brasier, *Applications of selected reaction monitoring (SRM)-mass spectrometry (MS) for quantitative measurement of signaling pathways.* Methods, 2013. **61**(3): p. 313-22.
229. Goyal, A., et al., *S100b as a prognostic biomarker in outcome prediction for patients with severe traumatic brain injury.* J Neurotrauma, 2013. **30**(11): p. 946-57.
230. Pilotto, A., et al., *Differences between plasma and CSF p-tau181 and p-tau231 in early Alzheimer's disease.* medRxiv, 2021: p. 2021.12.10.21267467.
231. Naskar, A., et al., *Fibrinogen and Complement Factor H Are Promising CSF Protein Biomarkers for Parkinson's Disease with Cognitive Impairment–A Proteomics–ELISA-Based Study.* ACS Chemical Neuroscience, 2022. **13**(7): p. 1030-1045.
232. Marques, T.M., et al., *Identification of cerebrospinal fluid biomarkers for parkinsonism using a proteomics approach.* NPJ Parkinsons Dis, 2021. **7**(1): p. 107.
233. Rotunno, M.S., et al., *Cerebrospinal fluid proteomics implicates the granin family in Parkinson's disease.* Sci Rep, 2020. **10**(1): p. 2479.
234. Karayel, Ö., et al., *Accurate MS-based Rab10 Phosphorylation Stoichiometry Determination as Readout for LRRK2 Activity in Parkinson's Disease.* Molecular and Cellular Proteomics, 2020. **19**(9): p. 1546-1560.
235. Barkovits, K., et al., *Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition.* Mol Cell Proteomics, 2020. **19**(1): p. 181-197.

236. Gobom, J., *Sample Preparation for Proteomic Analysis of Cerebrospinal Fluid*, in *Cerebrospinal Fluid Biomarkers*, C.E. Teunissen and H. Zetterberg, Editors. 2021, Springer US: New York, NY. p. 175-180.
237. Wilhelm, M., et al., *Deep learning boosts sensitivity of mass spectrometry-based immunopeptidomics.* Nature Communications, 2021. **12**(1): p. 3346.
238. Holden, S.K., et al., *Progression of MDS-UPDRS Scores Over Five Years in De Novo Parkinson Disease from the Parkinson's Progression Markers Initiative Cohort.* Mov Disord Clin Pract, 2018. **5**(1): p. 47-53.
239. Karayel, O., et al., *Proteome profiling of cerebrospinal fluid reveals biomarker candidates for Parkinson's disease.* Cell Rep Med, 2022. **3**(6): p. 100661.
240. Yamada, T., et al., *Increased concentration of C4d complement protein in the cerebrospinal fluids in progressive supranuclear palsy.* Acta Neurol Scand, 1994. **89**(1): p. 42-6.
241. Wang, Y., et al., *Complement 3 and factor h in human cerebrospinal fluid in Parkinson's disease, Alzheimer's disease, and multiple-system atrophy.* Am J Pathol, 2011. **178**(4): p. 1509-16.
242. Jabbari, E., H. Zetterberg, and H.R. Morris, *Tracking and predicting disease progression in progressive supranuclear palsy: CSF and blood biomarkers.* Journal of Neurology, Neurosurgery & Psychiatry, 2017. **88**(10): p. 883.
243. Svenningsson, P., *Corticobasal degeneration: advances in clinicopathology and biomarkers.* Curr Opin Neurol, 2019. **32**(4): p. 597-603.
244. Magdalinou, N.K., et al., *A panel of nine cerebrospinal fluid biomarkers may identify patients with atypical parkinsonian syndromes.* Journal of neurology, neurosurgery, and psychiatry, 2015. **86**(11): p. 1240-1247.
245. Quint, W.H., et al., *Bispecific Tau Antibodies with Additional Binding to C1q or Alpha-Synuclein.* J Alzheimers Dis, 2021.
246. Niimi, Y., et al., *Cerebrospinal Fluid Profiles in Parkinson's Disease: No Accumulation of Glucosylceramide, but Significant Downregulation of Active Complement C5 Fragment.* J Parkinsons Dis, 2021. **11**(1): p. 221-232.
247. Goldknopf, I.L., et al., *Complement C3c and related protein biomarkers in amyotrophic lateral sclerosis and Parkinson's disease.* Biochem Biophys Res Commun, 2006. **342**(4): p. 1034-9.
248. Alberio, T., et al., *Parkinson's disease plasma biomarkers: an automated literature analysis followed by experimental validation.* J Proteomics, 2013. **90**: p. 107-14.
249. Sun, C., et al., *Peripheral Humoral Immune Response Is Associated With the Non-motor Symptoms of Parkinson's Disease.* Front Neurosci, 2019. **13**: p. 1057.

250. Dufek, M., et al., *Serum inflammatory biomarkers in Parkinson's disease.* Parkinsonism Relat Disord, 2009. **15**(4): p. 318-20.

251. Chen, Y., et al., *The complement and immunoglobulin levels in NMO patients.* Neurol Sci, 2014. **35**(2): p. 215-20.