The proteome of neurofilament-containing heteroaggregates in blood as source of biomarkers for neurodegeneration

Rocco Adiutori

Submitted in partial fulfillment of the requirements of the Degree of Doctor of Philosophy

Academic supervisor: Dr. Andrea Malaspina

Industry supervisor: Dr. Ian Pike

I, Rocco Adiutori, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Details of collaboration and publications

Publication:

Adiutori R., Aarum J., Zubiri I., Bremang M., Jung S., Sheer D., Pike I., Malaspina A. The proteome of neurofilament-containing protein aggregates in blood. *Biochem Biophys Rep*. 2018 May 25; 14:168-177. doi: 10.1016/j.bbrep.2018.04.010.

Zubiri I., Lombardi V., Bremang M., Mitra V., Nardo G., Adiutori R., Lu CH., Leoni E., Yip P., Yildiz O., Ward M., Greensmith L., Bendotti C., Pike I., Malaspina A. Tissue-enhanced plasma proteomic analysis for disease stratification in amyotrophic lateral sclerosis. *Mol Neurodegener.* 2018 Nov 7; 13(1):60. doi: 10.1186/s13024-018-0292-2.

Collaborations:

All mass spectrometry data acquisition was performed in collaboration with Proteome Sciences plc, including bioinformatic analysis in Chapter5.

Transmission electron microscopy imaging was carried out in collaboration with Ms. Giulia Mastroianni, TEM facility manager at NanoVision Centre, School of Biological and Chemical Sciences, Queen Mary University of London.

«'O frati,' dissi, 'che per cento milia

perigli siete giunti a l'occidente,

a questa tanto picciola vigilia

d'i nostri sensi ch'è del rimanente

non vogliate negar l'esperïenza

di retro al sol, del mondo sanza gente.

Considerate la vostra semenza:

fatti non foste a viver come bruti,

ma per seguir virtute e canoscenza.'»

«'O brothers!' I began, 'woe to the west

'Through perils without number now have we reach'd;

'To this the short remaining watch, that yet

'Our senses have to wake, refuse not proof

'Of the unpeopled world, following the track

'Of Phoebus. Call to mind from whence ye sprang:

'Ye were not form'd to live the life of brutes,

'But virtue to pursue and knowledge high.'»

(from The Divine Comedy -- Inferno 26, By Dante Alghieri. English translation by The Rev. Henry Francis Cary, A.M.)

Acknowledgement

This work was supported by the Medical Research Council, MRC Industry Case Studentship - MR/M015882/1.

I would like to thank my primary supervisor Dr. Andrea Malaspina, who has been a great support in accomplishing this work. I would also like to thank my secondary supervisor Dr. Ian Pike, whose support and technical advices on the proteomics aspect of this work have been essential, and Prof. Gavin Giovannoni for supporting this project within the School.

I would also like to thank the entire Proteome Sciences, which kindly supported this work in all the aspects of the proteomics analysis. In particular, I want to thank Dr. Michael Bremang, who has been essential in all bioinformatics analysis, for his availability and kindness and Dr. Emanuela Leoni who was responsible for managing this project within Proteome Sciences.

The work on protein aggregates would not have been so quick without the collaboration with Dr. Johan Aarum and Prof. Denise Sheer. Their experience on protein aggregates was very important since the beginning for developing the work.

All imaging data presented in this thesis was acquired in collaboration with Giulia Mastroianni, NanoVision Centre. I would like to thank her for teaching me how to use the instrument, for time she spent analysing and discussing with me the micrographs.

Thanks also to the other member of the group, especially Dr. Irene Zubiri and Dr. Vittoria Lombardi who helped me interpreting the data, for all the fruitful scientific discussion and for their support and friendship outside the lab.

There have been hard times in this journey that I could overcome thanks to my friend, near and far. My heartfelt thanks to you all.

Finally, I would like to thank my family, which made this possible since the very beginning, and Chicca, for whom words cannot be enough.

Abstract

The lack of effective disease-modifying treatments and of means to achieve an early diagnosis strongly support the search for reliable neurochemical biomarkers in neurological conditions like Amyotrophic Lateral Sclerosis (ALS), a fatal neurodegenerative disorder. In ALS, the accumulation of protein aggregates containing neurofilaments (Nf), the building blocks of axons, leads to motor neuron death. Here we investigate the presence of neurofilaments (Nf) in protein aggregates in blood, studying a hallmark of neurodegeneration systemically. We explore the hypothesis that these circulating protein assemblies may function as biomarkers for neurodegeneration in accessible biofluids, which may have future application in clinical practice.

In this thesis, I developed a protocol based on ultracentrifugation for the enrichment of protein aggregates from blood and confirmed their enrichment in Nf. Using Mass Spectrometry (MS)-based proteomics, I have obtained data on the protein composition and functional relevance of Nf-Containing Hetero-aggregates (NCHs) in plasma samples from ALS patients, healthy controls (HC) and from ALS brains. I have then applied quantitative proteomics analysis using a TMTcalibrator™ workflow on plasma samples from ALS patients and HC using brain tissue as an internal calibrator. Additional experiments were undertaken to evaluate NCHs resistance to proteases digestion and to characterise the specific conformation of these macromolecular structures by Transmission Electron Microscopy (TEM). Our fluid-brain tissue investigation using a multi-modal approach suggests that NCHs may represent a systemic readout of biochemical changes identified in neurodegenerative brain pathology and suggest that they may acquire altered biochemical properties like protease resistance. The changes in NCHs identified in ALS compared to HC are a promising biological substrate for the future development of next generation biomarkers of neurodegeneration.

Table of contents

List of figures

List of tables

Abbreviations

ACN	Acetonitrale
AD	Alzheimer’s disease
ALS	Amyotrophic Lateral Sclerosis
ALSFRS-R	ALS functional rating scale revised
ALS-PPS	Amyotrophic lateral sclerosis-Pooled Plasma sample
AM	Ammonium Molybdate
Ambic	Ammoniumbicarbonate
BBB	Blood brain barrier
BOP	Brain Only Protocol
BP	Biological Processes
BPP	Brain-Plasma Protocol
bRP	Basic Reverse Phase
CalDIT	Calibrator Data Integration Tool
CB	Capture Buffer
CC	Cellular Components
CFs	Correction factors
CID	Collision-Induced Dissociation
CNS	Central Nervous System
CS	Chondroitin sulphate
CSF	Cerebrospinal Fluid
CSPGs	Chondroitin sulphate proteoglycans
CV%	Coefficient of variation

DA	Dopamine
DLS	Dynamic light scattering
DTT	Dithiothreitol
ECD	Electron-Capture Dissociation
ECL	Chemiluminescence substrate
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immune assays
ER	Endoplasmatic reticulum
ESI	Electrospray Ionisation
ETD	Electron-Transfer Dissociation
FA	Formic acid
fALS	Familial ALS
FAT	Functional Analysis Tool
FDR	False discovery rate
FeaST	Feature Selection Tool
G3P	Glycerol-3-phosphate
GAGs	Glycosaminoglycans
GO	Gene Ontology
GPCRs	G-protein-coupled receptors
HC	Healthy controls
HC-PPS	Healthy Controls-Pooled Plasma sample
HDL	High-density lipoproteins

HS	Heparan sulphate
HSPGs	Heparan sulphate proteoglycans
IAA	Iodoacetamide
IF	Intermediate Filaments
IgGs	Immunoglobulines G
IG-TEM	Immuno-gold Transmission Electron Microscopy
IQR	Inter-quartile range
iTRAQ	Isobaric tags for relative and absolute quantitation
KEGG	Kyoto Encyclopedia of Genes and Genomes
Knn	K-nearest neighbours
KS	Keratan sulphate
KSP	Lysine-serine-proline
LB4X	Loading Buffer 4X
LC	Liquid chromatography
LC-MS/MS	Liquid Chromatography coupled with Tandem Mass Spectrometry
LDL	Low-density lipoproteins
LF	Loading factor
LIMMA	Linear models for microarray
LMN	Lower motor neurons
logFC	Log2fold change
LPA	Lysophosphatidic acid
LPS	Lowest-probability subsequence
m/z	Mass-to-charge ratio

MetOH	Methanol
MF	Molecular Functions
MND	Motor neurone disease
MS	Mass Spectrometry
MS/MS	Tandem MS
MW	Molecular weight
NBB	Netherlands Brain Bank
NCHs	Neurofilament-Containing Hetero-aggregates
Nf	Neurofilaments
NfH	Neurofilament heavy
NfL	Neurofilament light
NfM	Neurofilament medium
NTA	Nanoparticles tracking analysis
ON	Overnight
PA	Phosphatidic acid
PAPA	Prion Aggregation Prediction Algorithm
PC	Phosphatidylcholine
PCA	Principal component analysis
PD	Proteome Discoverer 1.4
PE	Phosphatidylethanolamine
PGs	Proteoglycans
pI	Isoelectric point
PI	Phosphatidylinositol

PLAAC	Prion-like amino acid composition
PNS	Peripheral Nervous System
PPAR	Peroxisome proliferator-activated receptors
PRL	Progression Rate at Last Visit
PSMs	Peptide-Spectra Matches
PTMs	Post-Translational Modifications
PVDF	Polyvinylidene difluoride
RBPs	RNA Binding Proteins
RT	Room Temperature
sALS	Sporadic ALS
SDS	Sodium dodecyl sulphate
SEP	Seprion PAD-beads
SOD1	Superoxide dismutase 1
SR	Seprion Reagent
TB	Transfer Buffer
TBS	Tris-Buffered Saline
TEM	Transmission Electron Microscopy
TFA	Trifluoroacetic acid
TMT	Tandem Mass Tag
UA	Uranyl Acetate
UC	Ultracentrifugation
UMN	Upper motor neurons
VLDL	Very low-density lipoproteins

WB	Western Blot

1 Introduction

1.1 Neurodegenerative disorders: the example of Amyotrophic Lateral Sclerosis

The increasing prevalence of neurodegenerative diseases, in conjunction with the so-called silver tsunami of a rapidly aging population, is an alarming challenge to health services worldwide. The motor and cognitive impairment that neurodegeneration causes strikes right at the centre of our societies and of our ways of living: the burden on patients, family and carers is immense and the demographic as well as socio-economic scenario of a world facing a rapidly aging population with major neurological disabilities is rapidly becoming a reality. Prevention and treatment of neurodegenerative disorders is becoming a major health priority. However, the lack of biomarkers for early diagnosis and patient stratification is a limiting factor in the development of new clinical trials, along with the lack of novel therapeutic targets. The problem has both quantitative and qualitative connotations. Among conditions with a relatively slow progression, Parkinson's disease has an expected prevalence of 9 million globally by 2030, while Alzheimer's disease will affect between one third and one half of all individuals above the age 80 by 2050, with health expenditure predicted to reach 1 trillion dollars annually in the US alone (74,75). Motor neurone disease (MND)/amyotrophic lateral sclerosis (ALS) prevalence is expected to increase by 69% by 2040, reaching a global prevalence of about 380,000 (76). However, unlike neurodegenerative diseases with slow progression, ALS can have a dramatic development to end-stage disease in less than a year from disease onset, requiring expensive nutritional and ventilatory support (77) Families, shocked with bereavement, are left to piece their lives back together when MND strikes.

ALS is a fatal neurodegenerative disease with no effective treatment, affecting motor neurons in the brain and spinal cord and leading to death by respiratory paralysis in 3 to 5 years. ALS has an incidence of one or two new cases per year per 100,000, with males being more likely affected than females (78).

The disease clinical expression is unpredictable and can vary dramatically between patients as shown in Figure 1.1. However, two main groups can be

distinguished as bulbar onset, where bulbar motor neurons are affected from the outset, and limb onset, where somatic/spinal motor neurons are affected. Another aspect of the disease is the involvement of upper motor neurons (UMN) and lower motor neurons (LMN). UMN originate in the motor cortex of the brain and are connected to LMN through their axon in the spinal cord at cervical area where the LMN cell body is located and they are connected to the muscles through their axon. When neurodegeneration starts at the UMN level there are signs of muscle spasticity and brisk reflexes, while for LMN degeneration the signs are weakness, muscle atrophy and fasciculation (79). The diagnosis of ALS is normally based on evidence of involvement of both upper and lower motor neurons, in the same or in different anatomic areas.

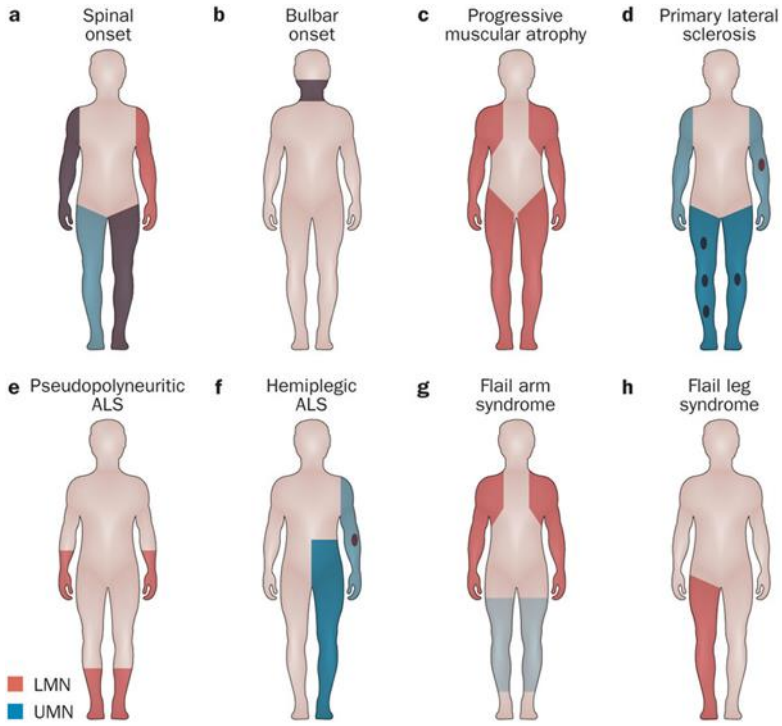


Figure 1.1. Different onset sites involving upper and lower MNs in ALS patients (79)**.**

This variety in disease phenotypes, length of survival and the way the neurodegeneration is detected in the brain and spinal cord, has generated a prion-like hypothesis for ALS, whereby the pathological protein aggregates in affected tissues are responsible for disease development and its spread (80,81). In fact, the hallmark for ALS, as with other neurodegenerative diseases, is the presence of protein aggregates in motor neuron cytoplasm, with TDP-43 as the most represented protein within these inclusions. Other proteins have been found to be involved in protein aggregation in ALS, such as NfH and Superoxide

dismutase 1 (SOD1) (82). Also, these inclusions within neuron bodies seem to be subjected to a wide range of chemical and enzymatic modifications (83,84).

About 10% of the overall ALS cases show a family history for the disease and are defined as familial ALS (fALS), while the remaining 90% are defined as sporadic (sALS). Among fALS only 68% of the cases are linked to genetic mutations and the same set of genes account for only 11% of the sALS (Figure 1.2) (85).

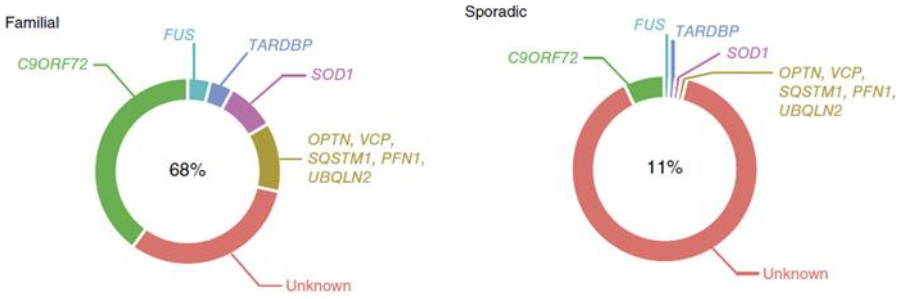


Figure 1.2. Genes involved in sporadic and familial ALS. Adapted from (85)**.**
The charts show relative percentages of cases with a known genetic cause for fALS and sALS.

There are different mechanisms proposed for ALS pathogenesis, including impaired proteostasis, disturbed RNA metabolism and RNA Binding Proteins (RBPs) transcription, defects in cytoskeletal and axon-transport, impaired DNA repair, defects in vesicle transport, excitotoxicity related to increased glutamate concentration at synaptic level and mitochondrial dysfunction (86). Although all these processes seem to intervene in ALS development, it is not yet clear which are causative, and which are consequences.

The diagnosis of ALS is primarily based on the symptoms and signs the physician observes in patients and on a series of tests to rule out other diseases, which can mimic ALS. This lengthy process results in a diagnostic delay of about 12 to 18 months from onset.

Methods to quantify the disease burden in ALS and the accumulation of neurological disability have been developed which include the ALS functional rating scale revised (ALSFRSR; range 1 to 48 where 48 is a healthy state and 1 the maximum level of neurological compromise) (87). To estimate rate of progression, for the purpose of our investigation we will use an established scoring method which is based on the calculation of the slope or loss of ALSFRSR points every month from disease onset to the last assessment

(Progression Rate at Last Visit – PRL; PRL > 1 fast progressors, PRL< 0.5 slow progressors).

Equally, a measure of the stage of development of the disease is provided by the El Escorial Criteria for diagnosis, which are based on the anatomic extension of the combination of upper and lower motor neuron signs (88).

1.2 Neurofilament proteins

Neurofilament (Nf) proteins are type IV Intermediate Filaments (IF), expressed in Central and Peripheral Nervous System (CNS, PNS), and constituted by three different proteins, Nf Light (NfL), Medium (NfM) and Heavy (NfH), encoded by three different genes. As other IF, they share the same structure with a head domain at N-terminus, a rod domain and a tail of different lengths at C-terminus (Figure 1.3).

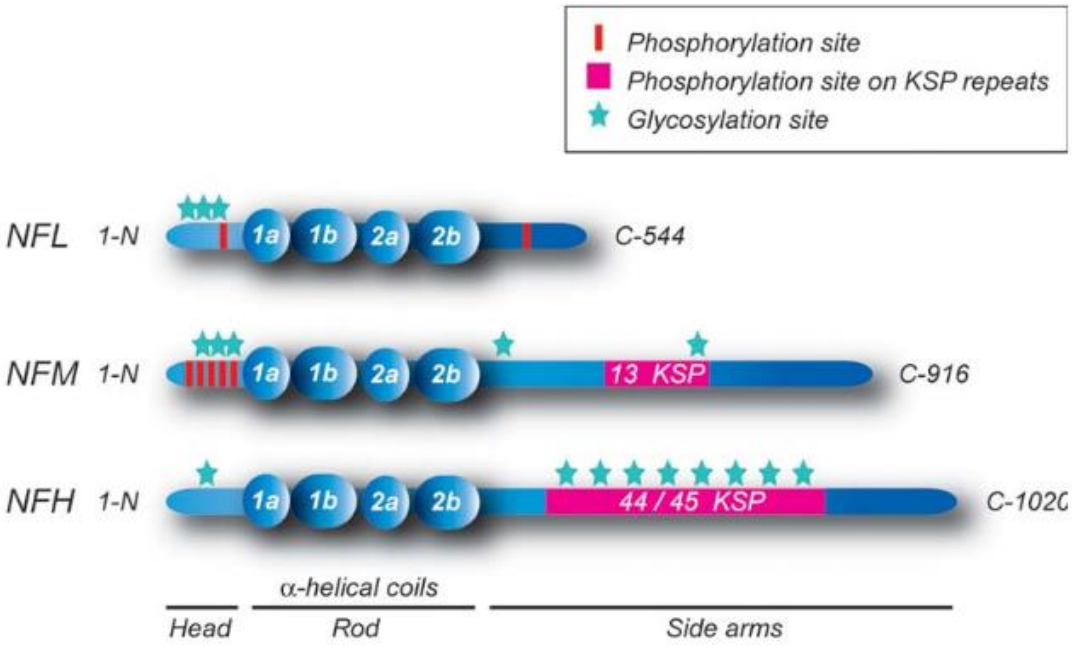


Figure 1.3. Schematic representation of neurofilaments (Nf) structure.

Neurofilament (Nf) Light (NfL; 543 amino acids (aa), 68 KDa), Medium (NfM; 916 aa, 160 KDa) and and Heavy (NfH; 1026 aa, 205 KDa). The cartoon highlights the head domain at the N-terminal of the protein, the rod domain, which is important during assembly of different Nf isoforms (divided in four segments), and the Tail domain at the C-terminal, which varies in length defining the three Nf isoforms. Tail and head are subjected to different types of post-translational modification and is also made of different numbers of lysine-serine-proline (KSP) repeats (13 in NfM and more than 40 in NfH) which are important for regulation of Nf function.

The Rod domain is an α-helical, coiled-coil rich region, that can be divided into 4 different domains (1a, 1b, 2a and 2b), also shared with others IF (1). This central domain is responsible for assembly with other Nf proteins through ionic interactions (2), firstly generating a dimer in a parallel orientation and then a two dimers complex in an anti-parallel orientation giving rise to a tetramer (3). This is the block that will form a filament unit, made of eight tetramers interacting with each other, while axial polymerization of filament units will form the actual filament, with Nf tail domains extending outward from the filament axis (4) (Figure 1.4).

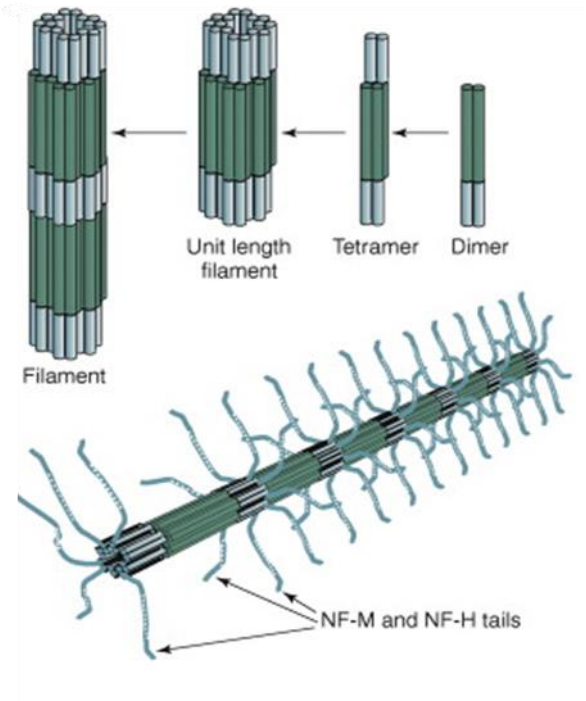


Figure 1.4. Assembly and formation of Nf.
In the filament, NfM and NfH tails radiate outward owing to the extensive charges from phosphorylated residues. Adapted from (5).

Nf are the most abundant cytoskeleton component in neurons. Together with Microfilaments and Microtubules are important constituents of neuronal structure (6). In addition to be a structural component of the cytoskeleton, Nf interact with other proteins (e.g. motor proteins), organelles, vesicles, mitochondria and plasma membrane (7–12). Nf are implicated in neuronal growth and in the maintenance of axonal caliber, playing a central role in axonal conduction properties (13,14).

During neuronal development, NfL is the first subunit to be expressed replacing peripherin (5), this process being followed by the expression of NfM and NfH and by a reduction of microtubules density (15–17). However, while NfL seems to be necessary in Nf assembly supporting axon and cytoskeleton development, NfM and NfH appear to be important in axonal caliber growth (18,19). Experiments performed in NfM and NfH knockout mice have shown that retaining either protein is enough to increase axon caliber during development and establish a good neuronal network, while NfH expression is required for *distal to proximal* development (20).

An important biochemical feature of Nf is the Post-Translational Modification (PTMs) of amino acids which affect their functional properties. Phosphorylation at serine and Threonine is the most commonly described PTM which is known to affect assembly and transport of Nf and increasingly implicated in the pathogenesis of neurodegenerative disorders (21,22). The main phosphorylation sites in Nf are the KSP repeats that are very abundant in the tail region, making NfM and NfH the isoforms with the highest level of expression of these PTMs. Phosphorylation does also occur in the head domain where it is thought to have a role in reducing Nf assembly in the cell body, allowing flow of Nf to the axon where they are mostly required (23–25). In fact, studies of Nf phosphorylation have shown that this PTM is higher in the axon and lower in cell bodies and dendrites and this may also be explained by cross-bridges formation between Nf subunits and Microtubules which determine axonal caliber (26–28). Ultimately, it is thought that the tight regulation of Nf phosphorylation in the head and tail domains is an important mechanism thought to protect neurons from abnormal Nf accumulation (29).

Another important PTM is Glycosylation, in particular O-GlcNAcylation at serine and threonine residues (in competition with phosphorylation) which can affect Nf assembly (30). Glycosylation seems to be an important regulator of Nf ubiquitination, whereby a small molecule called ubiquitin is attached to a protein which becomes inactivated and signals the protein-transport machinery to ferry the target protein to the proteasome for degradation (31). Beside ubiquitination, another mechanism proposed for Nf degradation is the activation of Calpain, a protein belonging to the family of calcium-dependent, non-lysosomal cysteine

proteases (proteolytic enzymes) expressed ubiquitously in mammals and many other organisms. Nf chemical cleavage by Calpain would occur as a consequence of increased intracellular calcium levels (32). However, it is important to point out that different mechanisms that are involved in Nf modification may have competing functions. For example, phosphorylation may act as a shield against protease-mediated degradation protecting Nf integrity (33). The ubiquitin proteasome system has laso been described in Nf degradation, while it is unclear if lysosome and autophagy are involved in Nf clearance (34–37). In the context of brain injury, for example, microdialysis studies in areas of neuronal injury have shown how the mechanical impact is followed by the activation of Pavlov's enterokinase in the human brain, a neuronal proteolytic pathway that produces products of NfH cleavage (NfH(476-986) and Nf(H476-1026)), which have been used as specific protein biomarkers applicable to in vivo monitoring of diffuse axonal injury and neuronal loss in traumatic brain injury (38).

Different bodies of experimental evidence have uncovered another important aspect of the behavior of Nf, and in particular of the relative composition of combined Nf isoforms in tissues across species. Comparison of the molar ratios of neurofilaments and Nf isoforms isolated from rat, bovine and human brain shows a variation of each polypeptide contribution to the filament structure, suggesting that the molar composition of the neurofilament triplet, may have specific molar stoichiometry in each species, both during the developmental stage and in brain maturity (39). Recently, Nf subunit stoichiometry calculations and Monte Carlo simulations of a coarse-grained Nf brush model have been used to look at whether the blood expression of NfL, NfM and NfH isoforms in a rapidly progressing and invariably fatal neurodegenerative condition, amyotrophic lateral sclerosis (ALS), changes significantly from the physiological Nf subunit stoichiometry of 7 : 3 : 2 (NfL:NfM:NfH;) (40). The analysis showed the relative over-expression of NfL over the other two subunits, which has been interpreted as an energy and time-saving option for motor neurons shifting protein expression from larger to smaller (cheaper) subunits, at little or no costs on a protein structural level, to compensate for increased energy demands seen in a condition of neurodegeneration.

1.3 Biomarkers of neurodegeneration and neurofilaments

There is an important unmet need in modern medicine: the lack of biomarkers for neurodegenerative disorders to be used in clinic for diagnosis and prognosis. Due to the fact that neurodegenerative disorders present a high level of clinical heterogeneity and may also share important feature at presentation, it is currently difficult to have timely and accurate diagnoses based solely on clinical observation and other paraclinical means of investigations (41). More importantly, it is still impossible to anticipate disease onset in order to establish early treatments or interventions that could arrest or slow down disease progression in individuals at risk of developing a neurodegenerative disorder (42). Another important indication for the development of informative disease biomarkers for neurodegeneration is the need for tools to establish prognosis, which could be employed in the clinical stratification of phenotypic variants of neurodegenerative disorders. This has been at the centre of relentless efforts in a prevalent neurodegenerative disorder like Alzheimer's disease (AD) where the aim is an improvement design of clinical trials with the identification of a disease stage and a proper tool to monitor treatment efficacy (43,44). Similarly, the importance and at the same time the difficulty in defining a panel of robust disease biomarkers with high diagnostic performance has emerged in a more rapidly progressing and clinical heterogeneous neurodegenerative condition like ALS (43).

The most informative biological fluid to study neurochemical markers of neurodegeneration is the Cerebrospinal Fluid (CSF) since it is near the affected tissue, it provides the brain and spine with nutrients and eliminates products of neurodestruction. There are several proposed biomarkers for neurodegeneration in CSF, both representative of axonal and synaptic loss including Nf, Tubulin, Actin, Tau, Amyloid precursor protein, α-synuclein and N-acetylaspartic acid (45). Nf are released in cerebrospinal fluid (CSF) and blood when neurons and axons degenerate, and changes of Nf levels are associated with the progression of several neurodegenerative disorders (46–48). However, there are drawbacks in the use of this fluid for biomarkers analysis. In particular, longitudinal sampling with invasive lumbar punctures would not be practical for monitoring disease in clinical practice, given the variety of problems that may arise with disease progression including frailty of advanced patients and the lack of collaboration of cognitively impaired individuals which may require extreme means of sedation

(49). More accessible biofluids like blood and urine may be more suitable for disease monitoring and for the requirement of frequent and serial sampling. It is now acknowledged that most neurodegenerative conditions would develop a compromised blood brain barrier (BBB), enabling the leakage of different size molecules linked to neurodegeneration into blood, making this fluid a suitable target for biomarkers discovery (50,51). Furthermore, it has also been recently established that meninges harbor brain-draining lymph vessels, a novel route through which by-product of brain pathology may end up been excreted in the peripheral circulation (52).

Nf have been proposed as good blood biomarker for neurodegeneration by several studies in different neurological disorders (46–48). Lu et al. have recently shown that the immunodetection of NfH, but not of medium and light chain, when tested in increasing dilutions, lacks linearity compared to Nf recombinant proteins, a phenomenon described as the hook effect (53). We believe that this non-linear dilution curve depends on Nf sequestration by immune-complexes and/or by larger molecular hetero-complexes which are disrupted by the dilution process. In the context of the work reported in this Thesis, the use of the suffix “hetero” to refer to the potential macromolecular formations that sequester Nf in circulation, is based on the assumption that the molecules partaking in these complexes are of different nature, including proteins like antibodies, but also lipids and eventually nucleic acids. It has been recently shown that the creation of Nf fragments following degradation by proteases and their release in body fluids has been proposed as a mechanism for the generation of autoantibodies in neurodegenerative disorders (54), an example of immune reactivity which is currently evaluated as a potential source of disease biomarkers. Sample dilution is likely to perturb the heterocomplexes homeostasis, changing Nf epitope exposure to detection antibodies hence generating non-predictable readings on immunodetection (53).

1.4 Neurofilaments and protein aggregation in tissues and fluids

The formation of assemblies of proteins that have lost their soluble state is a pathological hallmark of several neurodegenerative diseases (55,56). The confluence of proteins into aggregates may also occur physiologically as shown for the recruitment of RNA-binding proteins into stress granules (57–61). In cells,

the aggregation of misfolded proteins is usually kept in check by a quality control system, which operates through protein re-folding, autophagy and clearance by the proteasome (62,63). Extracellularly, a range of immune mediators may contribute to the clearance of misfolded proteins and of their aggregated forms (64). It is proposed that in biological fluids, aggregate formation reflects the propensity of proteins to assemble naturally or can be experimentally induced under conditions of stress (65–67). Depletion of albumin from human plasma, for example, leads to a significant increase in protein aggregation, particularly when heat and shear stress are applied (68). Recently the presence of soluble sodium dodecyl sulphate (SDS) resistant protein aggregates has been reported in plasma from older adults and in significantly lower levels in plasma from younger individuals (69). Loss of protein homeostasis and the increased rate of intra-cellular protein aggregation seem to be important hallmarks of aging (69). Therefore, aggregates found in circulation may originate from senescent cell that have lost their functional integrity. Equally, an age-related failure of the control of protein homeostasis (i.e., proteostasis) may condition an increase of aggregation-prone proteins in fluid state and the formation of aggregates (69). Aggregates, inclusion bodies or aggregosomes described in neurodegenerative diseases including amyotrophic lateral sclerosis (ALS) (70), Charcot-Marie-Tooth disease (71) and Parkinson's disease (72), contain Nf suggesting a role in neurodegeneration and a possible application in the development of biological readouts to help in the diagnosis and prognostic evaluation (46,47).

As reported in Chapter 3, we suggest that circulating Nf are also present in high molecular weight molecular complexes. Therefore, understanding Nf distribution between brain tissue and fluids, and between low order oligomers and higher order hetero-aggregates in biofluids has significant implications on their utility as biomarkers. In ALS Nf form heterogenous protein aggregates (73) and it is assumed that these are released, essentially intact, into the blood stream following cell death. The de-novo formation of circulating heteroaggregates due to a seeding effect of proteins like Nf in the fluid phase cannot be excluded.

Therefore, the relative representation in biological fluids of Nf isoform and of Nf-containing heteroaggregates (NCH) depends on the following: 1) different kinetics of interaction between Nf and other proteins and/or auto-antibodies

(46,53); 2) Nf conformation and/or post-translational modifications which condition their immunogenicity and propensity to aggregate (5); and 3) the presence in the same milieu of other pro-aggregating proteins which may drive the assembly of Nf-enriched aggregates.

NfH may partake, at a systemic level, in a similar process of assembly in large and potentially toxic protein aggregates which can be seen developing within degenerating neurons (46,47). Therefore, a better characterisation of circulating Nf-containing hetero-aggregates (NCH) will help unravel the real potential of NfH, an emerging strong biomarker of neurodegeneration.

1.5 MS-based proteomics for biomarkers discovery

An important factor in the study of biomarkers of neurodegeneration in blood is the availability of assays capable of detecting very low amounts of analyte/epitope. Firstly, the blood-brain barrier blocks the passage of degeneration by-products from brain/CSF into blood. Secondly, the blood volume is higher than CSF diluting down analytes availability. Moreover, the complex blood matrix could interfere with assays requiring extensive validation (45). With improvements in Mass Spectrometry (MS) instruments and bioinformatics tools, proteomics approaches for biomarker discovery now seem to be a promising alternative to immunoassay (89).

Proteomics aim is to identify, quantify and characterize proteins within a biological system. Different factors have enabled a rapid development of this discipline with the establishment of more sensitive and robust LC-MS/MS systems and the availability of larger gene and genome sequence database as well as the development of bioinformatics and data-mining science. A standard workflow for MS based Proteomics is shown in Figure 1.5.

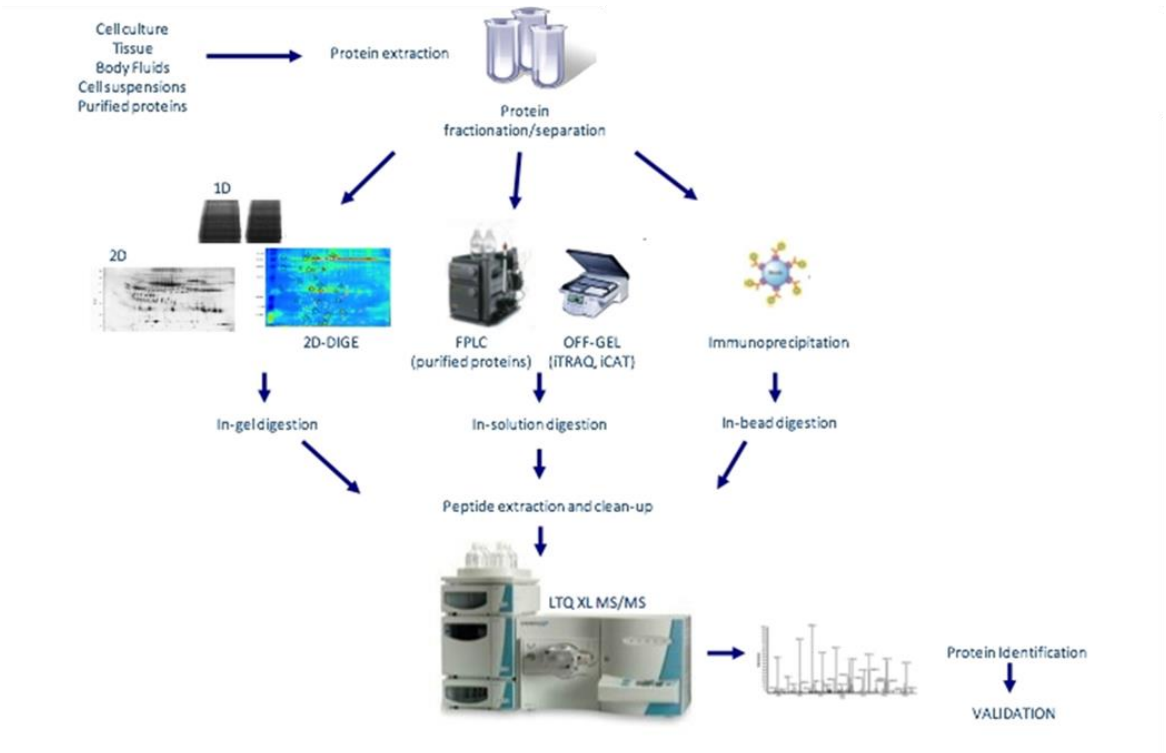


Figure 1.5. Standard workflow for a MS based Proteomics experiment with different strategies for sample preparation (gel based and gel free).

When analysing a complex biological fluid, tissue or cell lysate, a pre-analytical fractionation of the proteome of interest will favor the detection of a higher number of proteins and will enable the detection of low abundant proteins, whose detection is likely to be masked by more abundant proteins in samples with a wide dynamic ranges of protein concentration (90).

Bottom up proteomics is the most used approach nowadays in proteomic studies where the resulting peptides from an enzymatic digestion, typically trypsin, are analysed (opposite to the top down proteomic approaches that are based in the analysis of the intact proteins) (91). To increase peptide detection at MS level, high-resolution chromatography steps have been introduced upstream MS measurements to resolve the resulting complex mixture of peptides.

MS measures the mass-to-charge ratio (m/z) of ions in a gas-phase and operates with an ion source to produce gas-phase ions. Any mass spectrometer is composed by three basic elements: an ion source, a mass analyzer and an ion detector (Figure 1.6). Even if there are many different options and specific instrumentation based on the defined characteristic, at present one of the most

common setting in MS-based proteomics is Electrospray Ionisation (ESI) coupled with Orbitrap.

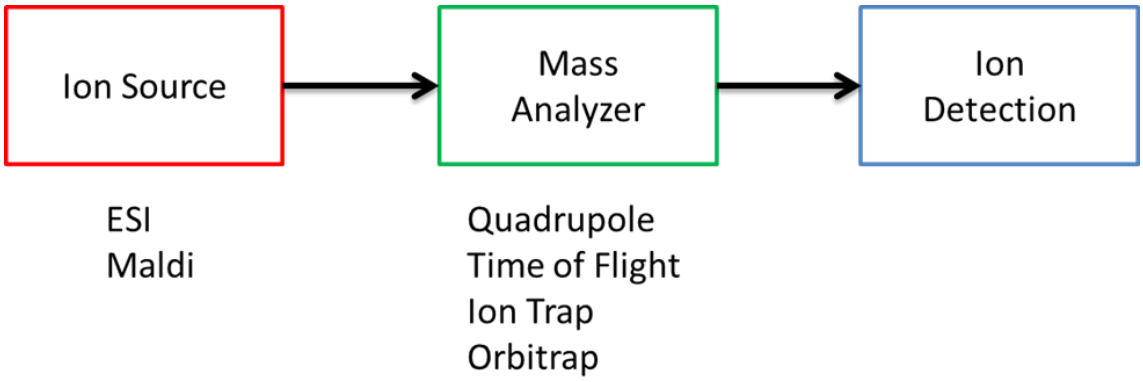


Figure 1.6. Schematic of MS with possible variants for ion source and mass analyser.

To improve peptide identification (through peptide sequencing) and characterization of PTMs, Tandem MS (MS/MS) is used. In this setting, the precursor peptide is fragmented and the resulting ions are measured. There are three fragmentation methods depending on the mass analyser and they produce different ion types: Collision-Induced Dissociation (CID), Electron-Capture Dissociation (ECD) and Electron-Transfer Dissociation (ETD). The generated ions, are then exploited to infer amino acid composition of these fragments with dedicated computational tools, producing Peptide-Spectra Matches (PSMs) that are crucial for protein identifications (92). Each PSM is given a score by the algorithm (there are several, such as SEQUEST, MASCOT, MS AMANDA) and the highest score assign the spectra to a protein. When all the spectra are inspected, the PSMs are then grouped to generate the final protein list.

Tandem MS allows also quantitative proteomics and many different techniques have been developed for this purpose (label free quantification, isotopic labelling, and isobaric labelling). In this report, we will focus on the use of isobaric mass tags, as the designed technique to be used in this study. Isobaric tags are a series of reagents that can be incorporated to the proteins/peptides and present the same mass in their original state. These tags, such as Isobaric tags for relative and absolute quantitation (iTRAQ) and Tandem Mass Tag (TMT), are composed of a mass reporter, which allows quantification, a mass normaliser, to maintain the same mass for the labelled precursor and amine reactive group to label proteins (Figure 1.7).

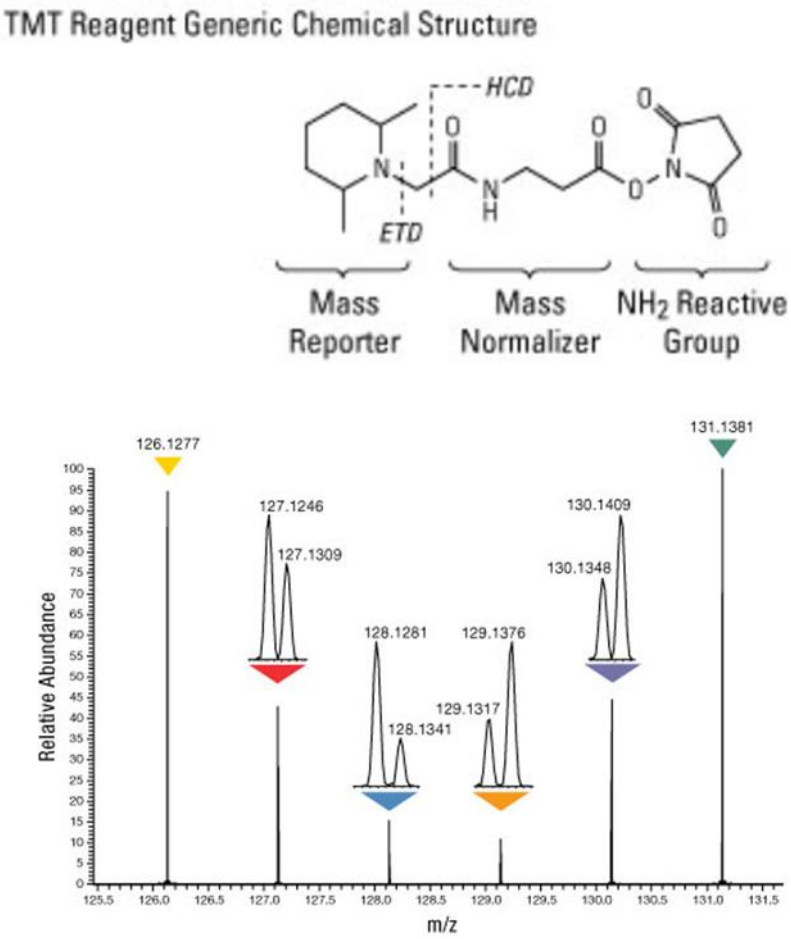


Figure 1.7. Structure and readout of an isobaric mass tag.

General structure of the TMT reagents with the different portions responsible for binding the tryptic peptides, keep the same weight for the first MS scan and discriminate the different samples in the MS2 scan, hence the possibility to quantify different samples in a single run.

After fragmentation, these tags will generate reporter ions that can be quantified, different samples will be labelled with different tags allowing the quantification of the relative abundance of a peptide in each labelled sample.

In conclusion, even if instrumentation and workflow in MS based proteomics is quite established, the choice for appropriate sample preparation and protein identification strategies is crucial. Low abundant proteins are difficult to detect and since in discovery experiments, as in the case of biomarker discovery projects, these proteins might be the most interesting it is important to isolate clean sub-fractions or deplete very abundant proteins.

In this study, MS analysis is used for the qualitative comparison of aggregates sub-proteome between ALS and healthy controls after enrichment of aggregates

from blood. A quantitative analysis will be performed using the TMTcalibrator™ (93) workflow with brain tissue as calibrant and purified blood aggregates as the 'fluid' channels, an approach that will potentially allow the detection of low abundant peptides in the analytical samples. With this setting, the calibrator proteins dominate the overall protein content within the injected fraction. The overall dominance of the calibrator signals in the TMT 10plex ensures the vast majority of peptides detected are from tissue-derived proteins. During MS acquisition, where a peptide is present in both the calibrator tissue and the body fluid, we obtain both the peptide sequence and quantitative TMT reporter ion signals from all 10 samples within the same MS/MS spectrum. This allows us to detect peptides that would otherwise be lower than the limit of detection for standard data-dependent acquisition, and results in a list of peptides that are common to the calibrator and the analytical samples. Moreover, to confirm NCH presence in the material extracted from plasma of patients and controls, we have also undertaken analysis by Immuno-gold Transmission Electron Microscopy (IG-TEM). This approach could also discriminate protein aggregates of different shape and nature establishing protein aggregates as possible biomarkers.

1.6 Aims of the Thesis

In light of recent observations on the presence of Nf-containing protein aggregates in bio-fluids, the work of this PhD studentship has centred around the development of methodologies for the isolation of such aggregates and their characterisation as disease biomarkers of neurodegeneration. This has been accomplished using biological samples from individuals with ALS, an aggressive and almost invariably fatal neurodegenerative disorder.

We hypothesise that neurofilaments may be released under both normal and pathological conditions as hetero-aggregates and that the content of these formations may differ between neurologically normal and diseased individuals. Hence defining the presence and content of NCH in normal individuals is a necessary step towards developing their utility as a new source of ALS biomarkers. Establishing protocols for NCH isolation and molecular characterization is therefore mandatory for any future use of NCH as disease biomarkers.

Based on the observation that neurofilaments-containing aggregates are hallmarks of motor neuron pathology, here we investigate the presence of these macromolecular formations in biological fluids, for the purpose of testing their potential as disease biomarkers in neurodegeneration.

The main aims of the experimental work of this Thesis are 1) to establish a robust and reproducible methodology for the separation of NCH from blood which provide suitable material for further characterization, 2) to employ complementary and new-generation forms of proteomics to study a disease-related molecular signature of the NCH and to compare the proteomes of NCH in biofluids with that of affected brains and 3) to provide an initial morphological investigation by electron microscopy and to test biochemical properties like protease resistance of the NCH obtained from healthy and affected individuals.

2 Materials and methods

2.1 Samples

2.1.1 Plasma

All plasma samples used in this study were obtained under ethical approval of the ALS biomarkers study (09/H0703/27). After consent, blood was collected from the study prticipant by venipuncture in EDTA tubes, processed within 1 hour by spinning at 3500 rpm for 10 minutes at 20 °C and later stored at -80 °C.

2.1.1.1 Sample pools

To develop and optimize the protocol for the extraction and analysis of Neurofilament-Containing Heteroaggregates (NCHs), plasma samples from different individuals were pooled together to minimise differences across samples and heterogeneity across individuals.

2.1.1.1.1 Healthy Controls (HC)

To test the method for isolation of Neurofilament-Containing Heteroaggregates (NCHs), a pool of plasma samples from healthy controls (Healthy Controls-Pooled Plasma sample, HC-PPS) was obtained using plasma from 6 Healthy Control (HC) individuals with known amount of plasma Neurofilament Heavy (NfH; 7.0 - 42.9 ng/ml). Data on NfH concentration was available from Dr. C. Lu previous work using an in-house ELISA (46). HC samples were selected with an age range (51.2-62.9 years) to be comparable with the age at sampling in patients with Amyotrophic Lateral Sclerosis (ALS), which were also sampled for the purpose of studying any disease-specific change in neurofilament-containng heteroaggergates (NCHs) (94).

2.1.1.1.1.1 Neurofilament expression in HC-PPS

To evaluate the presence of Neurofilaments (Nf)-containing high molecular weight protein aggregates in HC-PPS by Western Blot (WB; paragraph 2.3), plasma aliquots were filtered twice with Amicon filters 100K (Millipore), in order to enrich the content of proteins with high-molecular weight in the final test fluid. In this step, Nf Light isoform (NfL, ~70 kDa) was likely to be retained by the filter. As previously reported and detailed below, conditions known to solubilise plasma

aggregates were also employed and their effect on aggregates solubilisation tested (53). Pooled plasma samples aliquots prepared for NfH, NfM and NfL analysis were divided into three fractions and processed as follows: 1) pre-treatment with 0.5M urea and $Barb_2$EDTA buffer (60 mM Sodium Barbitone, 10 mM Barbital, 0.9 mM EDTA) for 1h at RT, 2) dilution 1:1 with $Barb_2$EDTA Buffer for 1h at RT and 3) left untreated at +4°C.

2.1.1.1.2 Amyotrophic Lateral Sclerosis (ALS) plasma samples

For the extraction of circulating Neurofilament-Containing Heteroaggregates (NCHs) from ALS patients, plasma samples from 6 ALS patients were pooled, including 3 from fast (ALS1-3) and 3 from slow (ALS4-6) progressing ALS patients, where progression rate of the disease was calculated as the decrease in the ALS Functional Rating Scale Revised (ALSFRSR) from disease onset to the time of sampling (48 – ALSFRSR at the time of sampling) divided for the time interval expressed in months (87,95). ALS patients with progression rate lower than 0.5 were classified as slow progressors, while patients with rate higher than 1.0 were classified as fast progressors. Plasma samples from fast and slow-progressing ALS patients were pooled (ALS-Pooled plasma sample ALS-PPS) to obtain a biological substrate representative of the phenotypic heterogeneity of the disease.

2.1.1.2 Individual plasma samples

Plasma samples from individual ALS patients and HC were also used in this study. Individual samples were chosen according to ethnicity, age at sampling, El-Escorial criteria for diagnosis (88), anatomic site of clinical onset, ALS Functional Rating Scale Revised (ALSFRS-R) (87), rate of progression (95) and volume of plasma available.

2.1.2 Brain tissue

Pre-central Gyrus brain tissue samples from two individuals affected by ALS (Brain1 and Brain2) were included in this project. The brain samples were obtained from The Netherlands Brain Bank (NBB), Netherlands Institute for Neuroscience, Amsterdam (open access: www.brainbank.nl). Brain had been collected from donors from whom written informed consent for a brain autopsy

and the use of the material and clinical information for research purposes had been obtained (NBB: under ethical permission 2009/148).

2.1.2.1 Brain tissue homogenization

Before aggregates enrichment (described below), brain tissue was homogenised as follows: a small portion of brain was scraped from frozen samples of Brain1 and Brain2 and then transferred into the same tube for homogenisation. Samples were mechanically fragmented and re-suspended on ice in homogenisation buffer (0.8 M NaCl, 1% Triton X-100, 0.1 M Ethylenediaminetetraacetic acid (EDTA), 0.01 M Tris at pH 7.4 and proteinase inhibitor cocktail (Sigma)). After mixing buffer and sample in a ratio 10:1 (v/w) the mixture was then sonicated at max power on ice for 5 minutes and centrifuged at 21000xg for 30 minutes at 4 °C in a bench-top centrifuge. Supernatant was then collected and pellet re-suspended in 10 volumes of homogenisation buffer on ice, again sonicated on ice at max power for 30 seconds and finally centrifuged at 21000xg for 30 minutes at 4 °C. Supernatants were finally pooled and used for aggregates enrichment and the pellet, which contains tissue debris, was resuspended in 8 M urea and stored at -80 °C.

2.2 Aggregates enrichment methods

2.2.1 Seprion PAD-beads (SEP) (Microsens Biotechnologies)

The Seprion PAD-beads (SEP) isolation method is based on the proprietary ligands effect in retaining protein aggregates (96). The proprietary protocol for aggregates extraction include the use of different reagents and of beads as reported below.

HC-PPS was mixed with 200 µl Capture Buffer (CB) and 100 µl of Seprion Reagent (SR), then 100 µl of SEP beads were added to HC-PPS, CB and SR in the same tube. The mixture was incubated for 30 minutes in agitation with a rotating mixer at Room Temperature (RT). The beads were washed with the proprietary buffer and collected in a magnetic rack for elution, using Loading Buffer 4X (LB4X) (Fisher Scientific), 50 mM Dithiothreitol (DTT) and heat treatment at 95°C for 10 minutes. The final SEP aggregate-containing fractions were not quantified for total protein concentration as the Loading Buffer was not compatible with standard techniques for protein quantification including Pierce™

BCA (ThermoFisher Scientific) or Bradford (Bio-Rad) protein assay kits. For Transmission Electron Microscopy (TEM) analysis of aggregates separated using Seprion, elution was carried out using 20 µl 0.1 M NaOH, 0.1% Triton X-100. Beads were removed and elute-neutralized with 1 M Tris-HCl.

2.2.2 Ultracentrifugation (UC)

Exploiting the poor solubility of protein aggregates, a sedimentation method by ultracentrifugation (UC) was developed to enrich protein aggregates from human tissues. Below, two alternative protocols are described that were applied to optimise aggregates extraction from plasma (HC-PPS and ALS-PPS) and brain samples. The best performing protocol was then applied for NCHs enrichment from individual plasma samples.

2.2.2.1 Brain-Plasma Protocol (BPP)

To identify the best conditions for enrichment of protein aggregates to be extracted from plasma and brain, different detergents (SDS, Triton X-100, Sarkosyl), at either 0.5% or 2% concentration as well as different NaCl concentrations (0.5 M, 1 M, 1.5 M) for *salting in* (a process which enhances solubility by increasing the ionic strength) were tested.

111 µl of 20% Triton X-100 were added to 1 ml of plasma (final concentration 2% Triton X-100) and incubated in agitation for 10 minutes at RT. For aggregates enrichment from Brain sample, the amount of 20% Triton X-100 solution added depended on the amount of tissue homogenised (paragraph 2.1.2.1), but the Triton X-100 final concentration was kept at 2%. Centrifugation at 21000xg for 15 minutes was then performed at RT and supernatant was collected for UC. At this stage, a 800 µl Sucrose Cushion (1 M sucrose, 50 mM Tris-HCl pH 7.4, 1 mM EDTA and 2% Triton X-100) was added into each UC tube (Open-Top/Self-Seal, PA, 8x51mm, Science Services GmbH) together with 500 µl of pre-cleared plasma onto the Cushion. After balancing the tubes, UC was undertaken using a Sorvall Discovery 100SE, equipped with a TFT 80.2 rotor for 2 hours at 50000 rpm. The rotor was pre-cooled to perform this step at 4°C. The supernatant was discarded by inverting the tube and the pellet re-suspended in PBS, 1.5 M NaCl and vortexed for 30 seconds for washing.

Following an additional UC step of 40 minutes, the final pellet was re-suspended in 100 µl PBS. After BPP optimisation, the solubilisation buffer was changed to 8 M urea, while 500 µl PBS was used for Transmission Electron Microscopy (TEM) applications. Samples for TEM re-suspended in PBS were subjected to an additional washing and UC step (50,000 rpm for 40 minutes, 4°C). The supernatant was discarded and the pellet re-suspended in 100 µl ddH_2O. After transferring the final sample into a clean tube, sonication on ice at max power for 5 minutes (Diogenode, Bioruptor) was perfomed to disrupt possible assembly produced by the strong g-force. Enriched fractions were stored at -80°C for further analysis.

The same protocol was applied for protein aggregates extraction from brain samples after tissue homogenisation as described in 2.1.2.1.

2.2.2.2 Brain Only Protocol (BOP)

The following protocol for brain aggregates enrichment is a modified version of that published by Greenberg and Davied in 1990 (97).

Homogenised brain was incubated at RT for 10 minutes with Triton X-100, to a final concentration of 1%, and with DTT, to a final concentration of 1 mM. This was followed by sonication at max power on ice for 30 seconds. Subsequently, using an UC tube, 500 µl of sample was loaded onto 800 µl of the BOP Sucrose Cushion1 (1 M sucrose, 50 mM Tris-HCl pH7.4, 1 mM EDTA, 1% Triton X-100 and 1 mM DTT) and the UC step was performed at 50000 rpm for 2 hours. The rotor was pre-cooled to perform this step at 4 °C.

The first part of the protocol was modified to make it suitable for aggregate extraction from plasma. Plasma samples were incubated at RT for 10 minutes with Triton X-100, to a final concentration of 1%, and with DTT to a final concentration of 1mM. This was followed by centrifugation at 21000xg for 15 minutes at RT. The supernatant was collected into a new tube and sonicated at maximum power for 30 seconds, on ice. At this stage, 800 µl of the BOP Sucrose Cushion1 were loaded into an UC tube and 500 µl of the plasma-processed sample were added onto the cushion. UC was then run at 50000 rpm for 2 hours at 4 °C, using a Sorvall Discovery 100SE equipped with a TFT 80.2 rotor. From

this step on the protocol was kept identical for plasma and brain aggregates enrichment by BOP.

After UC the supernatant was discarded. Each pellet was resuspended in 500 µl of 0.5% Triton X-100, 0.5% deoxycholic acid and 0.25% Sodium Dodecyl Sulphate (SDS) solution and sonicated at max power on ice for 30 seconds. This step was followed by UC at 50000 rpm for 2 hours at 4 °C using 800 µl of BOP Sucrose Cushion2 (1 M sucrose, 0.5% Triton X-100, 0.5% deoxycholic acid, 0.25% SDS). The pellet produced was then solubilised in 2% SDS, 8 M urea and stored at -80 °C.

2.3 Western blot analysis

Samples for Western Blot analysis (WB) analysis were prepared adding LB4X, DTT (final concentration of 50 mM) and water (where necessary) protein mixtures. Samples were heated for 10 minutes at 95°C before loading into a SDS-denaturing polyacrylamide gel for electrophoresis (SDS-PAGE). Reagents and conditions varied depending on the protein mix under investigation as reported below.

2.3.1 Gradient gels

Protein samples were loaded into a gradient gel (3-8% Tris-Acetate or 4-12% Bis-Tris; Fisher Scientific UK Ltd) along with the HiMark™ Pre-stained Protein Standard (Fisher Scientific UK Ltd). A current of 60V for 30 minutes was applied for electrophoresis and then switched to 120V for 1 hour (for 3-8% gels) or 1 hour and 15 minutes (when 4-12% gels were used). After electrophoresis, the gel was washed first with cold ddH_2O and then with 10% Methanol (MetOH) in 1X Transfer Buffer (TB). Proteins were then transferred onto a polyvinylidene difluoride (PVDF) membrane, previously activated with MetOH, at 45V on ice for 1 hour and 45 minutes with TB 1X, 10% MetOH. Subsequently, the membrane was blocked in blocking buffer (5% skimmed milk in Tris-Buffered Saline (TBS) 0.1% Tween-20 buffer (TBS-T 0.1%)) at RT for 1 hour. Overnight (ON) incubation was performed with primary antibody at 4 °C followed by incubation with secondary antibody (horseradish peroxidase (HRP) conjugated) for 1 hour at RT, with membrane washes between steps using TBS-T 0.1%. Lastly, the membrane was

incubated with enhanced chemiluminescence substrate (ECL) and visualised using a Chemi-Doc camera (Bio-Rad).

2.3.2 Fixed concentration gels

Protein samples were loaded into a gel with fixed concentration (10% Bis-Tris or 12% Bis-Tris (Fisher Scientific UK Ltd)) along with the SeeBlue™ Plus2 Pre-stained Protein Standard (Fisher Scientific UK Ltd). For gel electrophoresis, a current of 90V for 30 minutes was applied and then switched to 150V for 1 hour. After electrophoresis the gel was washed first with cold ddH_2O and then with Transfer Buffer (TB) 1X, 20% Methanol (MetOH). Proteins were then transferred onto a polyvinylidene difluoride (PVDF) membrane, pre-activated into MetOH, at 30V on ice for 2 hours with TB1X, 20% MetOH. Subsequently, the membrane was blocked in blocking buffer at RT for 1 hour. Overnight (ON) incubation was performed with primary antibody at 4 °C followed by incubation with secondary antibody for 1 hour at RT, with membrane washes between steps using TBS-T 0.1%. The membrane was then incubated with ECL and visualised using a Chemi-Doc camera (Bio-Rad).

2.3.3 Antibodies

The antibodies listed in the following table (Table 2.1), with relative dilutions and buffers, were used in this study for WB:

Primary antibodies	ID antibody	Species	Provider	Working condition
anti-Neurofilaments light (NfL)	clone EP675Y	rabbit	Millipore	1:1000 in blocking buffer
anti-Neurofilament Medium (NfM)	AB1987	rabbit	Millipore	1:1000 in blocking buffer
anti-Neurofilament heavy (NfH)	N4142	rabbit	Sigma-Aldrich	1:1000 in blocking buffer

Primary antibodies	ID antibody	Species	Provider	Working condition
anti-TAR DNA-binding protein 43 (TDP-43)	G400	rabbit	New England Biolabs	1:1000 in blocking buffer
anti-Ubiquitinylated proteins	clone FK1 | 04-262	mouse	Millipore	1:1000 in blocking buffer
anti-Fibromodulin (FMOD)	CSB-PA008755GA01HU	rabbit	Generon Ltd	1:1000 in blocking buffer
anti-Glypican-4 (GCP4)	LS-C375826	rabbit	Source BioScience UK	1:2000 in blocking buffer
anti-Byglican (BGN)	HPA003157	rabbit	Cambridge Bioscience	1:250 in TBS-T 0.1%, 5%BSA
anti-Cation-dependent mannose-6-phosphate receptor (M6PR)	ARP43519_T100	rabbit	Insight Biotechnology	1:500 in TBS-T 0.1%, 5%BSA
anti-Protein DJ-1 (PARK7)	HPA004190	rabbit	Cambridge Bioscience	1:250 in TBS-T 0.1%, 5%BSA
anti-Endophilin-B2 (SH3GLB2)	H00056904-B01P	mouse	Bio-Techne	1:500 in TBS-T 0.1%, 5%BSA

Primary antibodies	ID antibody	Species	Provider	Working condition
Secondary antibodies	ID antibody	Species	Provider	Working condition
anti-Rabbit IgG	P021702-2	swine	DAKO	1:50000 or 1:20000 in blocking buffer depending on the primary antibody
anti-Mouse IgG	A28177	goat	Thermo Fischer Scientific	1:20000 in blocking buffer

Table 2.1. Primary and secondary antibodies used for immunodetection by Western Blot. The table lists the antibodies (primary and secondary), the clone or product code (ID antibody), the host (species), the provider, the buffer and dilution conditions for each antibody.

2.3.4 Densitometry analysis

After membranes visualisation with the Chemi-Doc camera, images were acquired and processed by Image Lab (Bio-Rad). Band volumes were obtained using the “Volume Tools” function in Image Lab and “Adj. Vol. (Int)” values were used for relative quantification of the bands detected. Normalisation, quantification and statistical analysis were carried out using the Analysis ToolPak in a Microsoft Excel environment.

2.4 Transmission Electron Microscopy (TEM)

The TEM work including sample preparation and micrographs acquisition was carried out in collaboration with Giulia Mastroianni and Dr. Giacomina Bennardo, Queen Mary University of London - TEM facility managers.

Transmission electron microscopy (TEM) exploits the interaction between a specimen adsorbed onto a carbon coated grid and a beam of electrons passing through it to generate images of the material. The carbon coated grids are copper discs with a mesh which are covered with a very thin layer of carbon film. For the

visualisation of small particles, grids with holey carbon film can be used to improve micrographs resolution. In our experiments, to improve the contrast, hence the quality of the micrographs, negative staining with heavy metals was used. The heavy ions, generated from the heavy metal salts, increase the contrast by reflecting the electrons coming from the beam. The negative stain is performed by incubating the carbon grid carrying the adsorbed sample, with a solution containing the heavy metal salt. In our study, Ammonium Molibdate (AM) and Uranyl Acetate (UA) were tested as possible heavy metal salts for negative staining.

A glow-discharged 400 mesh grid coated with carbon or holey carbon was incubated with a droplet (≈10/20 µl depending on sample concentration) of aggregates-enriched sample and after a short time (final condition 10 seconds), the excess was removed by carefully touching to the grid edge with filter paper. Negative staining was obtained incubating the grid with a droplet (20 µl) of 2% w/v uranyl acetate (UA). After washing with ddH_2O, the grid was air-dried at room temperature and micrographs acquired by a JEOL JEM 1230 electron microscope.

2.4.1 Immunolabelling for immune-gold TEM (IG-TEM)

After sample loading onto the EM grid as described above, the same grid was blocked by incubation with 0.1% w/v BSA in PBS (blocking solution) for 10 minutes. Immunolabelling with Gold particles (IG-TEM) was performed by incubating the grid for 1 hour with the primary antibody which was washed out using blocking solution. The grid was then incubated with the gold beads-conjugated secondary antibody for 1 hour. The subsequent washing step was performed with PBS. In order to stabilize proteins interaction, the grid was incubated with 1% glutaraldehyde for 5 minutes and subsequently washed with ddH_2O. Surface was negatively stained with 1% w/v Uranyl Acetate (UA), washed with ddH_2O and air-dried to perform micrographs acquisition.

2.4.1.1 Antibodies

We report below the antibodies and conditions (dilutions and buffers) used for IG-TEM:

Primary antibodies	ID antibody	Species	Provider	Working condition
anti-Neurofilaments light (NfL)	clone EP675Y	rabbit	Millipor	1:100 in blocking solution
anti-Neurofilament Medium (NfM)	AB1987	rabbit	Millipore	1:100 in blocking solution
anti-Neurofilament heavy (NfH)	N4142	rabbit	Sigma-Aldrich	1:200 in blocking solution
anti-Apolipoprotein A-1 (APOA1)	MIA1402	mouse	Thermo Fisher Scientific	1:100 in blocking solution
Secondary antibodies	ID antibody	Species	Provider	Working condition
Anti-Rabbit IgG	G7402	goat	Sigma	1:2000 in blocking solution
Anti-Mouse IgG	G7652	goat	Sigma	1:2000 in blocking solution

Table 2.2. Primary and secondary antibodies employed for immunodetection by IG-TEM. Primary and secondary antibodies (clone or product code, ID antibody), species of origin, provider, dilution conditions and buffers in the IG-TEM experiments.

2.5 Shotgun Proteomics

To evaluate the protein composition of the enriched protein aggregate fractions, Liquid Chromatography coupled with Tandem Mass Spectrometry (LC-MS/MS) analysis was carried out. Two different sample preparation protocols were used. Data were acquired using an Orbitrap Velos Pro or Orbitrap Fusion Tribrid (Thermo Scientific).

2.5.1 In-gel trypsin digestion

In-gel trypsin digestion was applied to PPS samples to analyse NCHs composition. The protein samples were prepared for electrophoresis by adding LB4X, DTT and water and heat-denatured as described above. Samples were then loaded into a gel and electrophoresis was performed for protein separation as previously described. Gel staining was undertaken using Imperial^{TM} Protein Stain (ThermoFischer Scientific) for 1 hour at RT followed by overnight (ON) wash with ddH_2O. Gel bands were cut out and each band was washed with 100 µl 50:50 100 mM Ammoniumbicarbonate (Ambic): Acetonitrale (ACN) and kept at RT for 10 minutes. After discarding the solution, 100 µl ACN were added and kept for 5 minutes to obtain gel dehydration. Subsequently, the liquid was removed and gel pieces were dried in a vacuum centrifuge. Subsequently, gel bands were rehydrated for disulfide bonds reduction with 100 µl of 10 mM DTT in 100 mM Ambic and incubated for 30 minutes at 56 °C in a heating block while shaking. Thereafter, tubes were cooled down at RT, the excess of solution removed, and 100 µl of 55mM Iodoacetamide (IAA) in 100 mM Ambic was added for alkylation of the cysteine residues. Tubes were then incubated in the dark at RT for 30 minutes. Each fraction was again washed and dehydrated to obtain de-staining of the gel band pieces. After complete de-staining, each gel piece was rehydrated with a minimal volume of a solution 50 mM Ambic, 0.01 µg/µl Trypsin (V542A, Promega). Each tube was kept at 4 °C for 20 minutes to absorb the digestion enzyme (trypsin) and the excess was removed, while 50 mM Ambic was added to cover the gel pieces and keep them wet during enzyme cleavage. Each sample was then incubated ON at 37 °C for protein digestion. The following day, supernatants were collected into new tubes, gel pieces were dehydrated with ACN to extract the tryptic peptides. Then ACN supernatants were recovered and added to each of the previously collected supernatants. All samples collected were then freeze-dried in a vacuum centrifuge.

At this stage, each freeze-dried sample was re-suspended in 0.1% trifluoroacetic acid (TFA) and cleaned via zip-tip^{TM} (Millipore), a strong cation exchange resin in a plastic tip, before injection into the mass spectrometer. Sample cleaning and injection was carried out at Proteome Science facilities under close technical supervision. Qualitative analysis was performed coupling the mass spectrometer to an EASY-nLC 1000 (Proxeon) system. Samples were resuspended in 10 µL

of 2% ACN/0.1% formic acid (FA), and a 5 µl aliquot was injected onto a 75 µm × 2 cm nanoViper C18 Acclaim PepMap100 precolumn (3 µm particle size, 100 Å pore size; P/N 164705; Thermo Scientific) with an additional sample loading volume of 12 µl of 0.1% Formic Acid (FA) in H_2O using the Thermo Scientific EASY-nLC 1000 system. Peptides were separated at a flow rate of 250 nl/min and eluted from the column over a 60-minute gradient starting with 0.1% FA in ACN (5−30% over 50 minutes, then 30−80% between 50 and 54 minutes, continuing at 80% up to 58 minutes) through a 75 µm × 50 cm PepMap RSLC analytical column at 40 °C (2 µm particle size, 100 Å pore size; P/N ES803; Thermo). After Electrospray Ionisation, MS spectra ranging from 350 to 1800 m/z values were acquired in the Orbitrap at 30k resolution and the 20 most intense ions with a minimal required signal of 5,000 were subjected to MS/MS by rapid CID fragmentation in the ion trap. Protein identification was carried out using the Thermo Scientific Proteome Discoverer 1.4.

2.5.2 TMTcalibrator™

As previously mentioned, to study any disease-related change in the proteome of NCHs using TMTcalibrator™, plasma samples from a group of six individuals with ALS matched to a group of six healthy controls (HC) were selected. In the TMTcalibrator™ setting developed by Proteome Sciences (93), it is possible to combine in the same multiplex experiment up to ten samples (10plex) which function as analytical and calibrant samples. The same Tandem Mass Tags (TMT) label reagents are used to label uniformly all the samples in the study (Figure 2.1A). The use of TMT labelling and the combination of analytical samples with the calibrant sample in the same experiment allows 1) accurate quantification of the peptides and protein groups across samples and 2) an enhanced detection of peptides with low abundance in the analytical channels due to the prevalence of those peptides in the calibrant channels. To achieve this, analytical and calibrant samples had to be mixed and injected in a certain ratio in each 10plex experiment. The total amount of protein for the analytical channel had to be at 12X (2X for each analytical sample (n=6)), while the calibrant at 21X, divided in four different concentrations (1X, 4X, 6X, 10X) (Figure 2.1B). In addition, the calibrant channels worked as reference for the quantification of the analytical samples across channels in distinct 10plex experiments. For the characterisation of the NCHs proteome, plasma protein aggregate samples were loaded onto the

analytical channels, while a mixture 1:1 of Brain1 and Brain2 lysates was loaded at different concentrations onto the calibrant channels (Figure 2.1C).

Plasma aggregates were obtained as described in section 2.2.2.1. Brain lysates were processed dissolving brain in SysQuant Buffer (8M urea, phosphatase inhibitor (PhosSTOP™, Merck) and protease inhibitor (cOmplete™, Merck)). The mixtures with brain and SysQuant Buffer were vortexed for 5 minutes and then sonicated for additional 5 minutes at max power on ice to dissolve the tissue in the lyses buffer. Tubes were then centrifuged at 21000xg for 10 minutes at 4 °C to eliminate cell or tissue debris, and the supernatants were transferred into new tubes. The protein concentrations of Brain1 and Brain2 lysates were measure by Pierce™ BCA assay and the two lysates mixed 1:1 (w/v)/(w/v) (Calibrant).

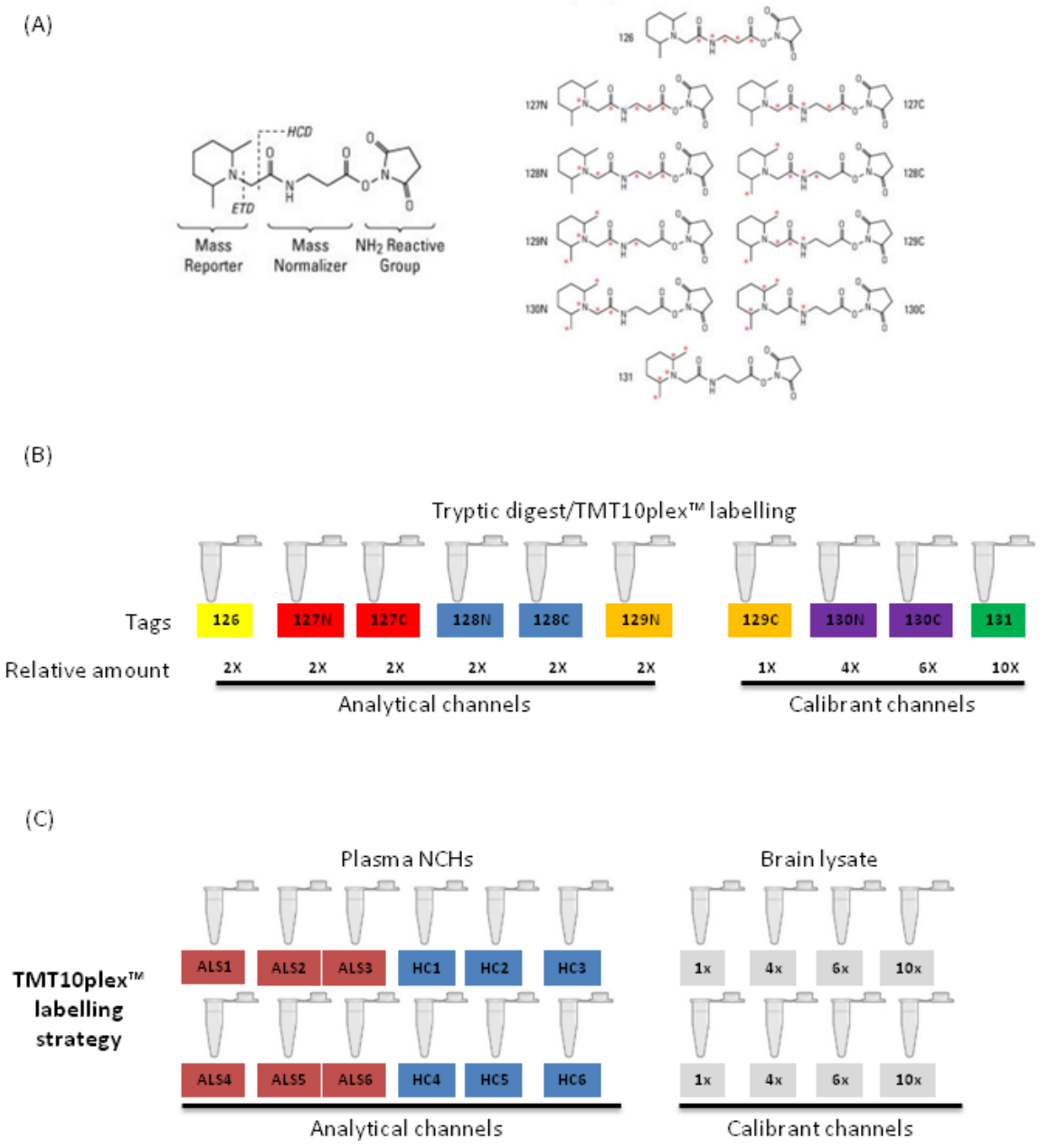


Figure 2.1. TMTcalibrator™ experimental design.

(A) Tandem Mass Tag (TMT) reagents with relative masses and isotope position. (B) General 10plex labelling layout after trypsin digestion of the samples; analytical samples and calilbrants are mixed in a specific ratio that enhances detection by LC-MS/MS of low abundant peptides in the analytical channels thanks to the high calibrant content. (C) Labelling strategy in two 10plexes LC-MS/MS runs which includes plasma neurofilament-containing hetero-aggregates (NCHs) from amyotrophic lateral sclerosis (ALS) patients and from healthy controls (HC) in the analytical channels (orange and blue colour codes) and a mixture (1:1) of brains (Precentral gyrus) lysates from two different ALS patients in the calibrant channels (grey colour code).

2.5.2.1 Sample preparation

TMT labelling, sample fractionation and injection was carried out at Proteome Science facility, Frankfurt, Germany. Sample fractionation and injection was performed under close technical supervision.

In order to establish the right analytical channels:calibrant ratio for each channel, 20 µg of total proteins were considered as 1X. Each sample was adjusted for the planned protein concentration and to reach the expected volume for TMT labelling adding SysQuant buffer.

0.1M DTT was added to each sample to a final concentration of 5 mM, followed by 25 minutes incubation at 56 °C in a heating block while shaking. After cooling, IAA was added to each tube to a final concentration of 14 mM and incubated at RT for 30 minutes in the dark. The fraction of IAA that did not react with DTT was quenched adding a surplus of DTT and using incubation in the dark for additional 15 minutes. Samples were then diluted 1:5 with 25 mM Tris-HCl pH 8.2 (to reach a concentration of urea equal 1.6 M), and trypsin (5 ng/µl) was added to a ratio of at least 1:100. Tubes were then incubated for digestion ON at 37 °C in a heating block while shaking. The day after, Trifluoroacetic acid (TFA) was added to the reaction mixtures (0.4% (v/v) of the total volume) to stop digestion, bringing the pH to < 2.0. Tubes were then centrifuged at 2500xg for 10 minutes and the pellets discarded, while the supernatants were collected into new tubes and processed with a 200 mg SepPak tC18 cartridge (Calibrant) and with 50 mg SepPak tC18 cartridge (analytical samples) for peptide desalting. After cartridges elution with ACN:H_2O (1:1), samples were collected in new tubes and the Calibrant volume was divided into four different aliquots with a 1:4:6:10 volume ratio (Calibrant channels), followed by a freeze-dry cycle. Calibrant aliquots contained enough sample for the two 10plexes.

Dried samples were re-solubilised in 120 µl KH_2PO_4 and TMT reagents were added combining a specific tag for each sample. TMT tags were previously resuspended in ACN to a concentration of 80 mM, and 40 µl of TMT reagents were added to each tube and incubated at RT for 1 hour. To stop the reactions, hydroxylamine was added in each tube to a final concentration of 0.25% (w/v) and incubated for 15 minutes.

The four Calibrant samples were pooled into a new vial and let to rest for 15 minutes. At the same time, six analytical samples for each 10plex were pooled into a new tube and incubated for 15 minutes. TFA in a 1:1 (v/v) ratio was added to each tube to bring the pH below 2 and H_2O to reduce total ACN concentration below 5%. Half of the Calibrant sample volume was mixed with the analytical samples of 10plex-1 while the other half was mixed with the analytical samples of 10plex-2. The samples containing the labelled peptides for the two 10plexes were desalted in 200 mg SepPak tC18 cartridge and the elute, containing now 660 ug of tryptic peptides, was freeze-dried before basic Reverse Phase (bRP) fractionation using Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (ThermoFisher Scientific). This step generated eight different fractions for each 10plex allowing higher peptide resolution and increasing the chance to detect (and quantify) low abundance peptides in the analytical samples using LC-MS/MS. All the eluted fractions underwent a further freeze-dried and desalting cycle.

LC-MS/MS analysis was performed in double-shot using a Thermo Scientific™ Orbitrap Fusion Tribrid (Thermo Scientific) mass spectrometer coupled to an EASY-nLC 1000 (Thermo Scientific) system. The 16 bRP fractions were resuspended in 2% ACN, 0.1% formic acid (FA), and then 12 µg from each was injected into a 75 µm × 2 cm nanoViper C18 Acclaim PepMap100 precolumn (3 µm particle size, 100 Å pore size; P/N 164705; Thermo Scientific). Peptides were separated at a flow rate of 250 nl/min and eluted from the column over a 5 hours gradient starting with 0.1% FA in ACN (5-30% ACN from 0 to 280min followed by 10min ramping up to 80% ACN) through a 75 µm × 50 cm PepMap RSLC analytical column at 40 °C (2 µm particle size, 100 Å pore size; P/N ES803; Thermo). After electrospray ionisation, MS spectra ranging from 350 to 1500 m/z values were acquired in the Orbitrap at 120 k resolution and the most intense ions with a minimal required signal of 10,000 were subjected to MS/MS by HCD fragmentation in the Orbitrap at 30 k resolution. Protein identification was carried out with Thermo Scientific Proteome Discoverer 1.4.

2.5.2.2 Bioinformatics

Bioinformatic analysis of the TMTcalibrator™ pipeline data was performed under close supervision of the Proteome Sciences bioinformaticians. MS analysis

generated 32 separate files that were submitted to Proteome Discoverer (PD) v1.4 (Thermo Scientific) using the Spectrum Files node, where spectra were filtered (700-10000 Da precursor mass, total intensity threshold 100, minimum peak count 8, S/N threshold (FT-only) 1.5). The SEQUEST-HT node was set up to search data against the human FASTA UniProtKB/Swiss-Prot database (Database Version: 9606_human_reviewed_170703.fasta) and assigned the best scored peptide sequences to the experimental spectra acquired at the level of second MS scan. The reporter ions quantifier node was set up to measure the raw intensity values for the TMT reporter ions (126, 127N, 127C, 128N, 128C, 129N, 129C, 130N, 130C, 131). The SEQUEST-HT search engine was programmed to search for tryptic peptides (with up to two missed cleavage sites allowed) and with static modifications of carbamidomethyl (C), TMT6plex (K) and TMT6plex (N-Term). Dynamic modifications were set to deamidation (N/Q) and oxidation (M). Precursor mass tolerance was set to 20ppm and fragment (b and y ion) mass tolerance to 0.02Da. Protein grouping was performed using Parsimony Principle option in the Protein Grouping area within PD (98). All raw intensity values were exported to tab delimited text files for further processing and filtering using softwares developed by Proteome Sciences. This included: 1) Calibrator Data Integration Tool (CalDIT) for peptide quantification and multiplexes merging; 2) Feature Selection Tool (FeaST) for quantitative comparison of different 10plex experiments and to evaluate significant macro-features of the data set; 3) Functional Analysis Tool (FAT) for identification of the enriched and regulated biological features of the identified proteome. The operations performed by the three bioinformatic tools are described below.

For all peptide-spectrum matches (PSMs), the reporter intensities in all channels were corrected for background and cross-talk between reporter ions at MS2. Then, reporter ion intensities of six plasma aggregates samples (126, 127N, 127C, 128N, 128C, 129N) for each TMT10plex were calibrated by a reference intensity computed from the average signal across the four brain lysate calibrant channels (129C, 130N, 130C, 131). The resulting PSM intensity values were log2-transformed. PSMs having no quantitative data were discarded. To generate a peptide matrix from all PSMs, the average of $\log_2$-ratios was calculated for each peptide sequence. Afterwards, peptide features with more than 34% missing quantitative values in either ALS or control experimental group were removed

from the data sets. The remaining missing quantitative values were replaced by values imputed by K-nearest neighbours (Knn) method (K=2) (99).

In order to reduce the bias introduced by two separateTMT10plexes in the same experiment, a batch effect removal procedure was applied. For this, a linear model was constructed to weigh the effect of each TMT10plex batch and channel. For protein-level analysis, expression values were computed by averaging (trimmed mean) peptides that were unique to the gene identifier. Quantitative data obtained through the bioinformatic pipeline were controlled using the median and the inter-quartile range (IQR) of peptide or protein distributions as quality controls. Both metrics are robust and stable to outliers and are therefore good for detecting extreme outliers. The median $\log_2$ expression and IQR were traced using control charts for all datasets at peptide and protein levels. A sample was considered as a strong outlier if either quality control metrics values were more than three standard deviations from the overall mean.

To study the variance structure of the data sets, principal component analysis (PCA) score and loading plots were generated. This method was used at different levels of data treatment to identify factors that influence the data variance (e.g. technical factors and/or sample inherent factors). Technical or experimental factors which influence the data were characterized in the scores and loading plots. The loadings plots allowed assessment of the magnitude and correlation of these effects.

To identify regulated peptides and proteins, a linear model using the R package “linear models for microarray (LIMMA)” was applied to perform a multivariate statistical analysis. Along with the peptide or protein quantitative values, (A) group (ALS/HC), (B) gender (M/F), (C) progression rate and (D) TMT batch (TMT01/TMT02) effect were included into the multivariate model to account for differences between individuals connected to these factors, and to enable a better assessment of the effect of each potential modifier. In order to include the HC group in the progression rate PCA, progression rate values close to 0 were randomly generated and imputed. Consequently, the influence of the experimental group “class” was analysed using the following linear model:

$$\text{logRatio} \approx \text{class} + \text{group} + \text{gender} + \text{progression rate} + \text{TMT batch}$$

Here $logRatio_{ij} = log_2(X_{ij}/X_{ref,j})$, logarithm of ratio of i^{th} sample relative to reference for feature j. Log_2fold change (logFC) and p-values were calculated for all features (peptides, proteins) used for statistical analysis. To consider a feature as “regulated”, the significance criterion α was set to 0.005 and a logFC threshold of $log_2(2)$ for statistical analysis at peptide level, while α was set to 0.05 and a logFC threshold of $log_2(1.5)$ for statistical analysis at protein level. Multiple testing corrections were performed using the Benjamini-Hochberg procedure, which provides a false discovery rate (FDR) for every statistical result. PCAs and Volcano plots were generated including the data of the significant features to visualize the results of the statistical analysis.

The functional analysis was performed to identify the biological processes driving the main differences between the classes of interest (ALS/HC). Two methods for analysis were selected: 1) significance of enrichment, evaluated by the Fisher exact test and 2) significance of regulation, evaluated by the Mann-Whitney U test. Functional terms used in this analysis included: Gene Ontology (GO) Biological Processes (BP), GO Molecular Functions (MF), GO Cellular Components (CC), Protein Domains/Families, Biological Pathways, microRNA substrates, and sequence Motifs. Human-specific annotations were extracted from a range of publicly-available data repositories. For the enrichment analysis, a two-sided p-value was generated by the Fisher exact test and the Benjamini-Hochberg method was used for multiple test correction. A minimum of two matched identifiers (e.g. gene names) was required for each peptide and terms with an adjusted p- value < 0.3 were considered significant. For the regulation analysis, a two-sided p-value was generated by the Mann Whitney U test and the Benjamini-Hochberg method was used for multiple test correction. In this procedure, expression values for all terms were compared to a background composed of 1000 randomly-sampled expression values that did not overlap with the term. A minimum of three matched identifiers (e.g. gene names) was required. Terms with an adjusted p- value < 0.3 were considered significant. All functional results were visualised in volcano plots (enrichment or median fold change vs. adjusted p-value).

2.6 Protease resistance assay

In recent studies, proteins involved in the pathogenesis of ALS have been investigated to evaluate the potential of a prion-like behavior, taking into account the pattern of propagation and the involvement of fibrillar protein aggregates in the disease progression (81,100). One of the main features of prion proteins is their resistance to protease enzyme digestion. Therefore, we have looked at whether NCHs present a resistance to protease digestion similarly to prion proteins and set up digestion protocols of aggregates enriched fractions with different proteases. Four different proteases were tested: Trypsin (V542A, Promega), α-Chymotrypsin (Chymotrypsin from now) (C4129, Sigma), Calpain (208712, Millipore) and Enterokinase (11334115001, Roche).

After plasma aggregates enrichment as described in section 2.2.2.1, each sample was re-suspended in the buffer (50 µl) recommended for each protease enzyme (PBS for Trypsin, 100mM Tris HCl for Chymotrypsin, 50mM Hepes, 30mM NaCl for Calpain and 50mM Tris HCl for Enterokinase), and 5 µl of 0.5 M DTT was added prior to proceeding with sonication on ice for 5 minutes. Protein assemblies produced by the strong g-force applied during UC steps, were disrupted and disulfide bonds reduced to enhance cleavage sites accessibility. In this way it was possible to avoid any protease resistance related to artefacts of the proposed digestion protocol. This was specifically important for Trypsin, as changes in the dynamic of aggregates digestion would have affected downstream LC-MS/MS analysis. Each enzyme was then added into the digestion reaction tube in a ratio 1:20 protease:total protein. In addition, 5.5 µl of 0.1 M $CaCl_2$ were added to Chymotrypsin and Calpain reaction tubes to activate the enzymes as indicated by the manufacturers. The reactions were subsequently incubated ON at 37°C to allow digestion. The day after, the reactions were stopped adding LB4X, DTT and by heating at 95°C for 10 minutes. Digested samples were then stored at -80°C until used for analysis by SDS-PAGE, followed by Zinc staining (Life Technologies), and NfH immunodetection, by Western Blot as described in 2.3.1).

2.6.1 NCHs profile analysis by SDS-PAGE

To evaluate differences between ALS and HC groups of NCHs protease digestion profiles, ALS and HC digestionproducts were separated and visualised in a SDS-

PAGE gel stained with Zinc. Six gels were ran each containing ALS and HC samples before and after digestion and this process was repeated for all protease enzymes. Images were then acquired by Chemi-Doc Camera.

For intra and inter-gel control, NCHs undigested profiles were used as reference to define digestion paterns and total protein content, while protease digestion and the undigested NCHs profiles were compared across ALS and HC different samples. The images acquired were processed by ImageJ to draw lanes boundaries and pre-selecetd lanes were plotted through the Analyze\Gels\Plot lanes function. Band intensities were obtained quantifying the area beneath the peaks within each lane. All NCHs bands, from digested and undigested samples, were also normalised using the Marker (HiMark™ Pre-stained Protein Standard) as reference. This allowed comparison of bands at the same MW in different samples across gels. Marker intensities were also used to generate correction factors to adjust loading differences across samples. Statistical evaluation was carried out in a Microsoft Excel Analysis ToolPak environment using a “t-Test: Two-Sample Assuming Unequal Variances” function with data analysis package in Excel2016.

2.6.2 NCHs protease digestion profile analysis by western blot against NfH

WB for NfH detection was performed on protease digested samples as described in section 2.6. The blots were acquired with a Chemi-Doc Camera and analysed by Image Lab (Bio-Rad). Band volumes were obtained using the “Volume Tools” function in Image Lab and “Adj. Vol. (Int)” values were used for relative quantification of the bands. Normalisation, quantification and statistical analysis were carried out in a Microsoft Excel Analysis ToolPak environment using a “t-Test: Two-Sample Assuming Unequal Variances” function with data analysis package in Excel2016.

2.6.3 Brain aggregates protease digestion

The same approach described above to study protein aggregates resistance to protease digestion was applied to NCHs enriched from a mixture of Brain1 and Brain2 samples, which were previously homogenized as described in section 2.1.2.1. The results obtained were qualitatively compared to the results obtained from plasma NCHs.

3 Protocol development for the enrichment of Neurofilaments-containing heteroaggregates (NCHs)

3.1 Introduction and aims

Neurofilaments (Nf) have been already described as part of assemblies of proteins and other molecules in brain (aggregates) and also to form aggregates in blood. Inclusion of Nf in large molecular complexes in bio-fluids may act as a confounder in attempts to obtain measurements of Nf levels for biomedical applications, such as in assays to establish diagnosis and prognosis of neurological conditions. NF detection may not be accurate, particularly with immunodetection which is based on antibody recognition of epitopes which may be masked by the assemblement of different proteins (46,53,101). In addition, it has been shown that protein aggregates are present in plasma at higher concentration in older age, suggesting that the biological processes responsible for protein aggregates formation may become more prominent with ageing. Ageing is undoubtedly one of the main risk factor for neurodegenerative diseases like Amyotrophic Lateral Sclerosis (ALS), Alzheimer's disease, Parkinson's disease and Huntington's disease, neurological conditions which have been linked pathogenically to the abnormal behaviour of disordered proteins and their propensity to aggregate (69).

In this chapter, I develop a protocol for separation from plasma and enrichment of Neurofilament-containing Heteroaggregates (NCHs). Development and optimization of the protocol has followed a two-step process. In the first instance, I use Seprion PAD-Beads (SEP) and a sedimentation-based approach by ultracentrifugation (UC) for NCHs enrichment from a Healthy Controls-Pooled Plasma Sample (HC-PPS). This extraction method is defined as Brain-Plasma Protocol (BPP) (Paragraph 2.2.2.1), when the same methodology is applied to the extraction of NCHs from brain. Secondly, I compare two UC-based extraction protocols here defined as BPP and Brain Only Protocol (BOP) (Paragraph 2.2.2.2), for NCHs enrichment from plasma of Healthy Controls Pooled Plasma Sample (HC-PPS), Amyotrophic Lateral Sclerosis-Pooled Plasma Sample (ALS-PPS) and brain tissue (a mixture of two brain tissue samples here called Brain1 and Brain2). In this Chapter, the methodology of NCHs separation is discussed

while qualitative differences in NCHs between ALS and HC are detailed in Chapter 4.

3.2 Neurofilaments expression in plasma

A pool of plasma samples was used to establish a method for enrichment of protein aggregates from blood, to confirm the presence of Nf proteins in these aggregates and to study their composition in health and disease in order to support any application in biomarkers discovery (Table 3.1). For the purpose of methodological development and to test different conditions for the final enrichment protocol, we have opted for samples pooling in order to reduce any variability across samples. The Healthy Control-Pooled Plasma Sample (HC-PPS) was obtained merging plasma samples from six different healthy individuals with a known concentration of neurofilament heavy chain (NfH) as reported in Table 1.

Sample ID	Gender	Ethnicity	Age at sampling (years)	NfH (ng/ml)
HC1-PPS	F	Caucasian	54.08 – 54.75	42.9
HC2-PPS	F	Caucasian	60.94	15.4
HC3-PPS	M	Caucasian	61.25 – 61.77	7.0
HC4-PPS	M	Caucasian	59.55 – 61.77	20.3
HC5-PPS	F	Caucasian	62.89	30.8
HC6-PPS	M	Caucasian	51.25 – 52.28	24.8

Table 3.1. List of plasma samples used for the Healthy Control-Pooled Plasma Samples (PPS) and demographics of donors.

For healthy controls (HC) 1, 3, 4 and 6, we report the age of the donors when samples were collected (more than one sample was needed to achieve the requested volume).

Sample ID: Identification code

Gender: M= Male; F= Female

Age at sampling: age of the patient at blood sampling (age-range shown if more than one time point sample was used)

NfH (neurofilament heavy; ng/ml): level of NfH are expressed in ng/ml

Initially the presence of Neurofilament (Nf) proteins in HC-PPS was evaluated by Western Blot (WB). Before gel-electrophoresis, two rounds of ultrafiltration with a 100 KDa cut-off filter (Amicon ultra100K, Millipore) were performed in order to

concentrate Nf proteins and increase their detection by WB. Filtration allowed retention of all molecules of molecular weight higher than 100 KDa. After Nf concentration by ultrafiltration, the samples obtained were divided into three different aliquots and treated as previously described by Lu et al. (53): one aliquot was incubated for 1 hour at RT with 0.5 M urea in $Barb_2EDTA$ buffer, a second one was diluted 1:1 with $Barb_2EDTA$ buffer and kept for 1 hour at RT, while the third aliquot was kept at 4 °C. WB was performed as described in Paragraph 2.3, but the transfer was carried out onto a nitrocellulose membrane instead of a PVDF membrane (Figure 3.1). Treatment with $Barb_2EDTA$ and urea has an effect in dissolving protein aggregates, hence allowing an assessment of the proportion of Nf retained within these formations.

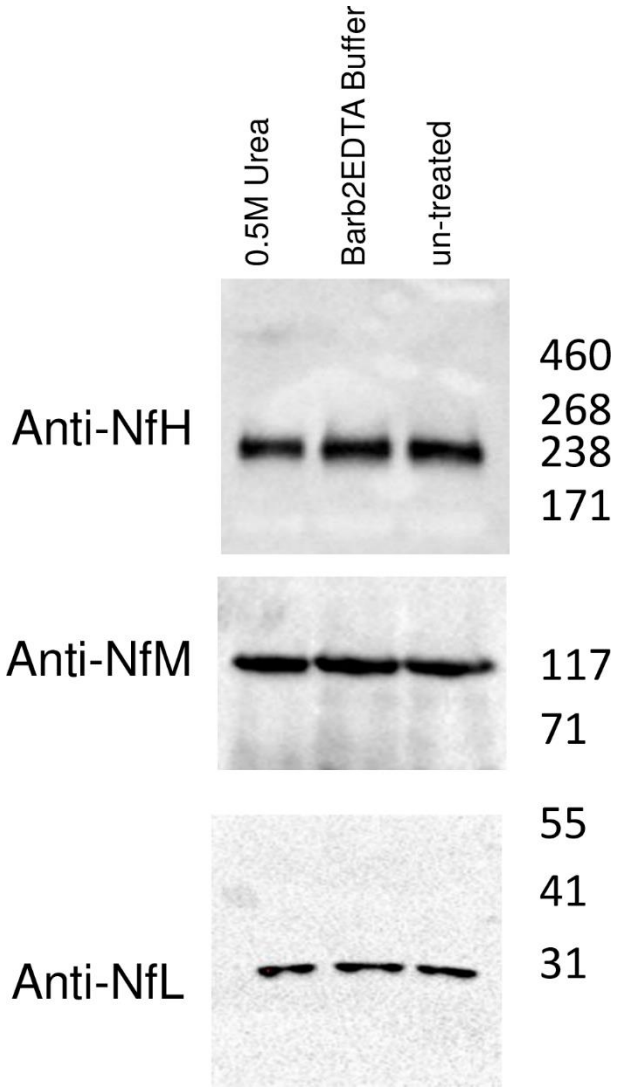


Figure 3.1. Western blot analysis of neurofilament proteins in healthy controls pooled plasma samples (HC-PPS).

Western blot analysis of pooled plasma sample (PPS) using anti-neurofilament-High (NfH), anti-Medium (NfM) and anti-Light (NfL) antibodies after filtration with Amicon 100 filters (100 kDa molecular cut-off). Lanes contain in order from left to right: PPS subjected to pre-treatment with 0.5 M urea and Barb2EDTA buffer for 1 h at RT, PPS diluted 1:1 with Barb2EDTA Buffer for 1 h at RT and untreated PPS kept at + 4 °C. The blots show the three neurofilament proteins, with NfL appearing at approx. 30 kDa, NfM at 117 kDa and NfH at 238 kDa. No bands with molecular weight higher than the expected sizes, indicative of stable Nf-containing aggregates, were detected.

Presence of Nf proteins was confirmed in HC-PPS. Neurofilament Heavy (NfH) and Medium (NfM) were detected at the expected molecular weight (MW; 238 KDa and 117 KDa respectively), while NfL was detected at 30 KDa (expected MW 70 KDa). NfL detection at a lower molecular weight may be related to the release by the SDS detergent used in the WB procedure of an NfL fragment from protein complexes retained by the filter. At the same time, there was no detection of SDS-stable immunoreactive bands above 460 KDa, with Nf bands intensity and MW appearing comparable across the three treatment conditions. As the presence of a NfL band at 30 KDa after ultrafiltration with a 100 KDa cut-off filter may have been related to this fragment being part of high-molecular weight assembly like a protein aggregates, we have further investigated the presence of high molecular weights aggregates in plasma using the reported enrichment protocols.

3.3 NCHs enrichment by Seprion PAD-beads

The first enrichment method tested was based on a ligand, called Seprion, capable of interacting with protein aggregates (96). In this study, we have used magnetic beads coated with the Seprion ligand (Seprion PAD-beads; SEP). To evaluate optimal binding conditions, different volumes of samples and reagents were tested in the final mix (Table 3.2). After SEP enrichment, each condition was evaluated by western blot (WB) for Neurofilament proteins (Nf) detection (Figure 3.2).

Seprion Enrichment Condition / Reagent	1	2	3	4	5	6	7
Sample (HC-PPS) [µl]	800	400	400	400	400	400	200
CB [µl]	200	400	200	200	400	400	200
Water [µl]	0	0	200	200	0	0	400
SR [µl]	100	200	100	100	100	100	100
Beads [µl]	100	200	200	100	200	100	100

Table 3.2. Conditions tested for optimization of plasma aggregates enrichment by SEP.

CB: Capture Buffer; SR: Seprion Reagent.

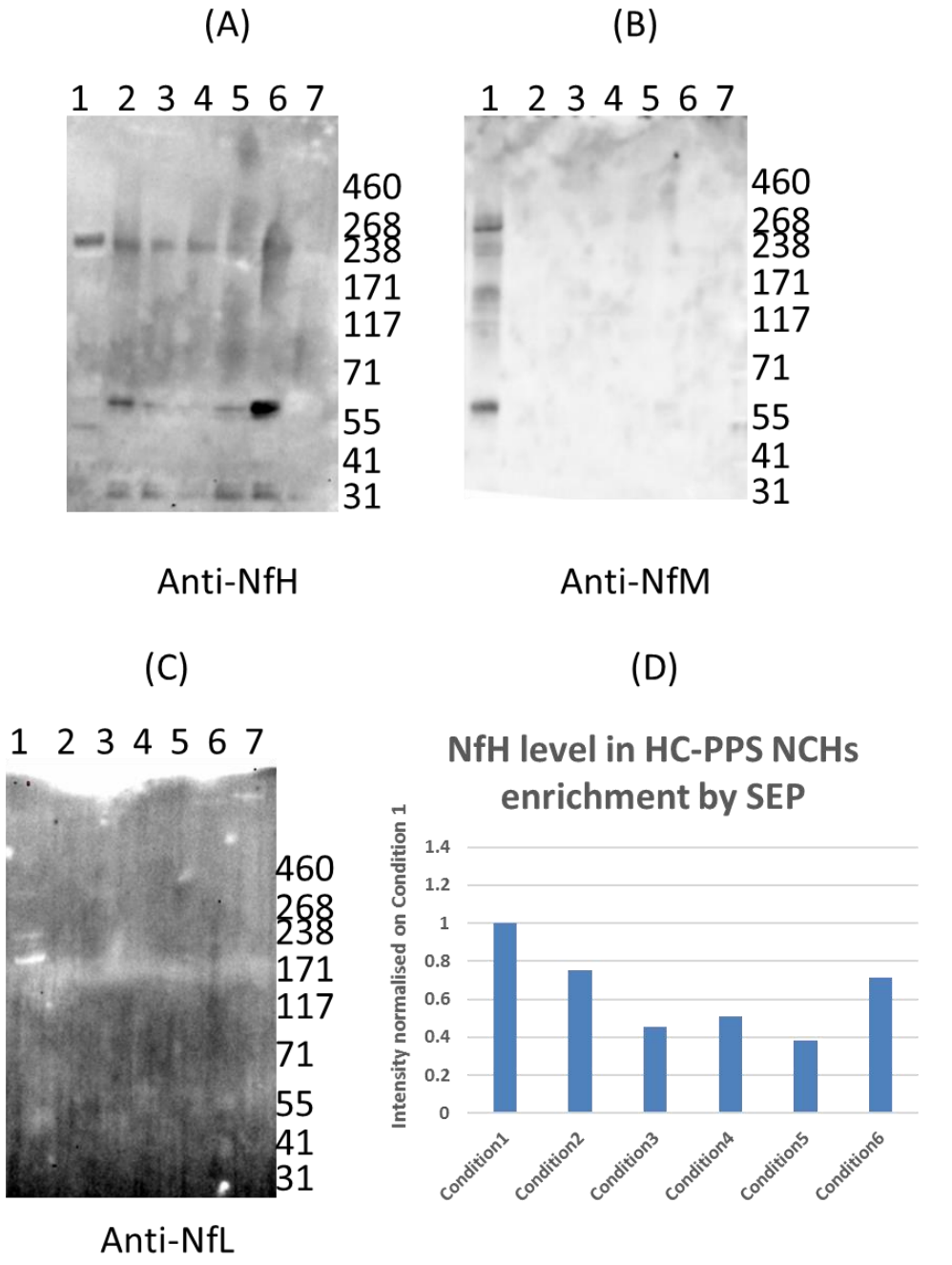


Figure 3.2. Nf in HC-PPS after aggregates enrichment with Seprion PAD-beads SEP. Western Blot (WB) analysis for the detection of (A) neurofilament heavy chain (NfH), (B) neurofilament medium chain (NfM) and (C) neurofilament light chain (NfL) in the healthy controls pooled plasma samples (HC-PPS) after SEP enrichment. The lanes labelled from 1 to 7 corresponded to HC-PPS sample enriched following conditions listed in Table 3.2. The chart showed in (D) represents the different amount NfH in (A) normalizing the intensities on condition 1 intensity.

It was possible to detect NfH at the expected molecular weight (MW) in all SEP-enriched HC-PPS samples, with the exception of condition 7 (Table 3.2). From

this blot, it appeared that the amount of NfH detected depended on the sample volume used for SEP. In fact, condition 7 where the smaller amount of HC-PPS was loaded (200 µl) showed no detectable NfH, while condition 1 with the highest amount of HC-PPS (800 µl) showed NfH expression of about double the intensity compared to the next condition tested (2; 400 µl) (Figure 3.2D). None of the NfH bands were above 268 KDa, the expected molecular weight for this protein. The correlation between volume of sample for SEP extraction and WB signal was also confirmed for NfM detection, where condition 1 only showed NfM expression. Bands at variable molecular weights were identified when WB was probed for NfM detection, with the most intense band detected at 55 KDa, possibly a fragment of the full-size protein released from the aggregates recovered by SEP.

Both for NfH and for NfM, the multiple bands detected could be related to different Nf peptides. NfL was not detected in any of the sample volumes used for SEP extraction.

3.4 NCHs enrichment by ultracentrifugation: Brain-Plasma Protocol (BPP)

Methods based on sedimentation (e.g. ultracentrifugation, UC) have been developed to separate macromolecular protein aggregates in solution which may be an unwanted by-product in biopharmaceutical formulations, as well as of small particles like exosomes and viruses (102–106). We have applied the same principle and developed a UC-based method to separate and enrich protein aggregates from plasma.

The rationale of this protocol is based on increasing the solubilisation of all the possible particles present in plasma by adding a detergent at high concentration and exploiting the high density of the residual, strong insoluble particles by sedimentation at high speed with a cushion of sucrose 1 M, which functions as a density cut-off (i.e. 1.127 g/ml). In addition, to further the enrichment of strong insoluble aggregates an additional *salting in* solubilisation step was included in the final protocol.

To identify the best conditions of enrichment by BPP, different detergents (SDS, Triton X-100, Sarkosyl) at either 0.5% or 2% final concentration and different salt (NaCl) concentrations (0.5M, 1M, 1.5M) for salting in were tested (Table 3.3).

Detergents and NaCl concentrations	Final concentration (µg/µl)
1 ml PPS + 25 µl SDS 20% (final concentration 0.5% SDS)	**0.30978**
1 ml PPS + 111 µl SDS 20% (final concentration 2.0% SDS)	**0.28813***
1 ml PPS + 25 µl Triton X-100 20% (final concentration 0.5% Triton X-100)	**0.86983**
1 ml PPS + 111 µl Triton X-100 20% (final concentration 2.0% Triton X-100)	**0.75391**
1 ml PPS + 25 µl Sarkosyl 20% (final concentration 0.5% Sarkosyl)	**0.66941**
1 ml PPS + 111 µl Sarkosyl 20% (final concentration 2.0% Sarkosyl)	**0.53673**
1 ml PPS + 111 µl NaCl 5 M (final concentration 0.5M NaCl)	**1.22388**
1 ml PPS + 250 µl NaCl 5 M (final concentration 1.0M NaCl)	**0.90196**
1 ml PPS + 430 µl NaCl 5 M (final concentration 1.5M NaCl)	**0.6338**

Table 3.3. Conditions of detergents and salt (NaCl) concentration: final protein concentration following enrichment of protein aggregates from HC-PPS using BPP.
Detergents (green colour-code) and salt (grey colour-code) conditions used for enrichment. *= SDS interference during protein quantification with Pierce BCA Protein Assay Kit.

The best detergent and salt condition for enrichment was based on Nf detection, both in its soluble form and within high MW forms (Figure 3.3). In fact, during the UC step performed at 4 °C, the SDS precipitated as powder at the bottom, especially the sample treated with 2% SDS that was not included in the western blot. The sample treated with 0.5% SDS showed a different pattern compared to the other treatment options with higher expression of NfH and NfM (Figure 3.3). It is therefore possible that the SDS concentration is the main factor in the enrichment process affecting aggregates solubilisation. Moreover, the NfM profile in the reported treatment condition was similar to the NfM profile showed by SEP enrichment in WB (Figure 3.2), suggesting that SDS detergent pre-treatment increased enrichment of protein aggregates carrying the NfM which is normally detected at 117 KDa (Figure 3.3). Pre-treatment of samples with a final

concentration of 2% Triton X-100, UC with 1 M sucrose cushion and washing with 1.5 M NaCl was associated with an overall higher amount of NfL, NfM and NfH as shown using WB (Figure 3.3). Under these conditions, NfH was consistently present at high MW (significantly above 460 KDa), while NfM and NfL both migrated at the same MW (117 KDa and 30 KDa respectively) reported for Nf detection in untreated plasma Figure 3.1. The detection of high MW, SDS-stable, NfH-containing bands in the enriched fractions was in line with the presence of circulating NfH-containing protein aggregates. We have therefore incorporated these conditions in the final BPP enrichment protocol (Figure 3.3).

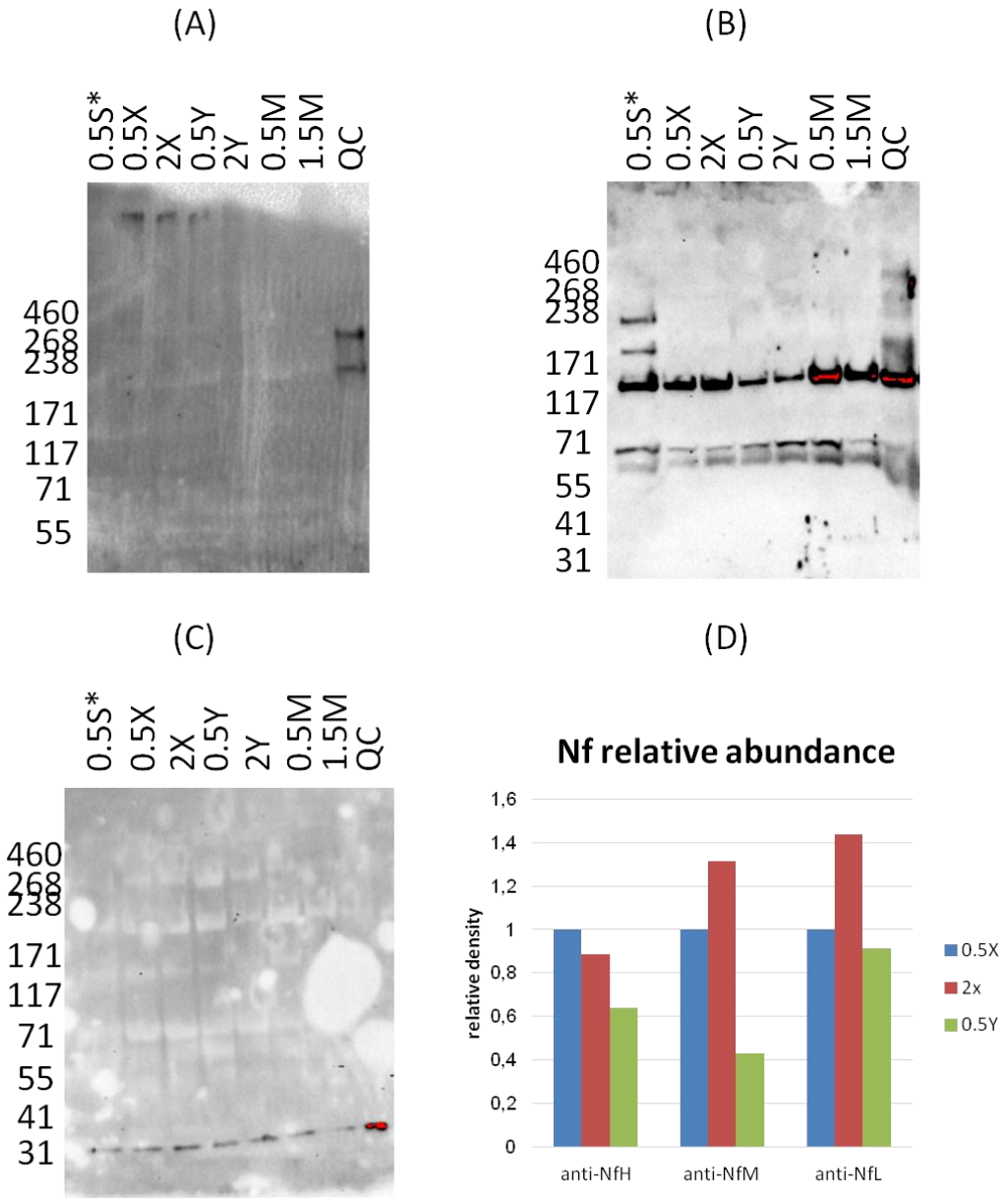


Figure 3.3. WB analysis of Nf isoforms using fractions generated by ultracentrifugation and pre-treatment conditions listed in Table 3.

11 µg of total proteins in each treatment sample were loaded to evaluate the best enrichment method, based on the level of Nf isofom detected. (A) Anti-NfH WB showing presence of high MW bands (above the expected NfH MW) for the 0.5X, 2X and 0.5Y conditions. (B) Anti NfM WB showing bands at 117 KDa MW for all condition tested and (C) anti-NfL WB showing no bands with MW higher than 31 KDa. *: 0.5S WB was associated with residual powder contamination, likely be due to poor SDS solubility at 4°C. (D) Abundance (relative density) of NfH, NfM and NfL in the WB experiments testing the three Nf isoform enrichment (A, B, C). The 2X condition showed the highest overall Nf content for all three Nf proteins. Data were normalised using the 0.5X bands density. QC: quality control, bovine neurofilaments (NfL, UmanDiagnostics; NfM and NfH, Stratech Scientific Limited). 0.5S: 0.5% SDS; 0.5X: 0.5% Triton X-100; 2X: 2% Triton X-100; 0.5Y: 0.5% Sarkosyl; 2Y: 2% Sarkosyl; 0.5M: 0.5M NaCl; 1.5M: 1.5M NaCl; QC: bovine Neurofilaments.

3.5 Comparison between SEP and BPP methods

The comparison of WB analysis of aggregates enrichment using the SEP and BPP methods showed higher expression of Nf species including high molecular weight NfH-containing aggregates when using the BPP enrichment protocol (Figure 3.1 and Figure 3.3). The BPP enrichment method identified NfL (not visible after SEP enrichment) and also high MW bands for NfH suggesting higher specificity in isolating macromolecular formations with this method. Further analysis by Liquid Chromatography coupled to tandem Mass Spectrometry (LC-MS/MS) and transmission electron microscopy (TEM) were carried out to characterize the enrichment effect offered by the two methodologies and to obtain preliminary data on NCHs morphology as well as composition.

3.5.1 LC-MS/MS

After aggregates extraction from HC-PPS using the SEP and BPP methods, the two enriched fractions were run into a gel to allow in-gel trypsin digestion as described in Material and methods (Paragraph 2.5.1). Before cutting the fractions for proteomic analysis, the gel was stained with Imperial Blue Comassie to identify the electrophoretic protein profiles generated using the two methods (Figure 3.4A). Based on these profiles, each of the two lanes was cut into fifteen fractions. Fractions were then digested with trypsin and subjected to LC-MS/MS analysis to study the composition of the aggregates extracted using the SEP and the BPP protocols (Figure 3.4A).

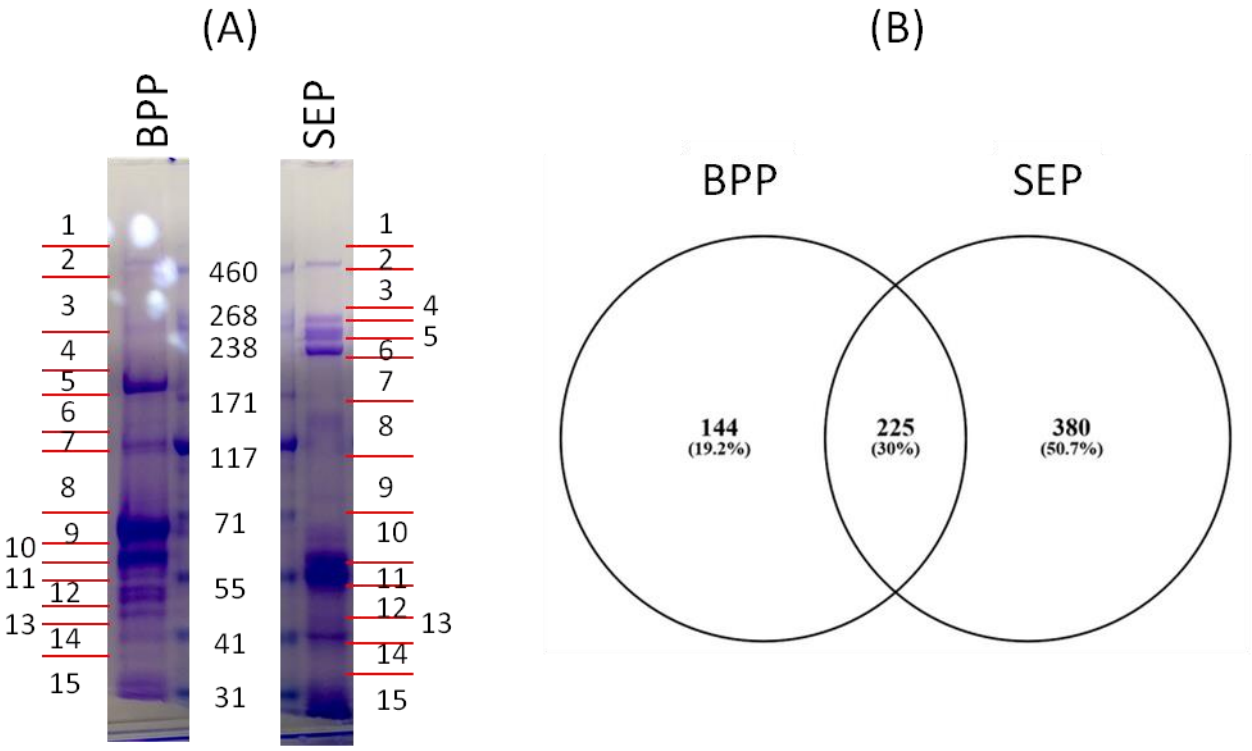


Figure 3.4. Aggregates fractions obtained using BPP and SEP methods resolved in SDS-PAGE and numbers of protein groups identified after LC-MS/MS analysis the fractions obtained using the two extraction methods.

(A) SDS-PAGE showing gel bands cut (fractionation) and subjected to in-gel Trypsin digestion for LC-MS/MS (a total of 15 bands for both BPP and SEP). (B) Protein Identification performed with Proteome Discoverer 1.4 (PD) generated 369 and 605 proteins unique to BPP and SEP respectively; 225 were shared by the two extraction methods (filter criteria for selection: a minimum of 2 peptides per protein and a peptide count only in the top scored proteins). LC-MS/MS: Liquid Chromatography coupled with tandem Mass Spectrometry; SEP: Seprion PAD-Beads separation; BPP: Brain-Plasma Protocol.

The mass spectrometry (MS) analysis identified 369 and 605 protein groups for the sample enriched by BPP and SEP respectively. However, Nf proteins were not detected in the two samples, possibly due to the presence of complex Post-Translational Modifications (PTMs) and to the low abundance of Nf (see also Paragraph 3.8). 30% of the overall detected proteins were observed in both samples (Figure 3.4B). The higher number of proteins identified in the SEP-enriched sample was not due to uneven sample loading: SEP and BPP samples had the same number of spectra acquired in the second Mass Spectrometer (MS2) and a comparable identification rate (ID rate), calculated as the number of Peptide-Spectrum Matches (PSMs) over the total number of spectra acquired

(Table 3.4). The higher Coverage% and number of PSMs per protein group in BPP (Table 3.4) suggested that this extraction method was more selective in the enrichment of specific proteins. The BPP and SEP shared protein pool had higher protein coverage and PSMs counts than those seen in either of the individual enrichment method, suggesting a higher abundance of these proteins in plasma (Table 3.4). Functional analysis of the three protein lists (Figure 3.4B) to define Genetic Ontology (GO) terms for Cellular Component (CC) was undertaken interrogating the PANTHER database (http://www.pantherdb.org/) (107), and KEGG pathway enrichment, using the Webgestalt web tool (http://webgestalt.org) (108).

Fraction	Total spectra	Total PSMs	ID rate%	Protein groups	Coverage% /protein group (mean)	#Unique Peptides /protein group (median)	#Peptides /protein group (median)	#PSMs /protein group (median)
BPP	215993	35779	16.56	371	31.2%	5.0	6.0	12.0
SEP	210875	34865	16.53	609	25.6%	5.0	6.0	8.0
BPP (shared)				209	34.7%	7.0	10.0	28.0
SEP (shared)				209	35.3%	8.0	10.0	21.0

Table 3.4. Liquid Chromatography coupled with tandem Mass Spectrometry (LC-MS/MS) analysis of the enriched aggregate fractions obtained using SEP and BPP: main features and protein groups defined by Proteome Discoverer 1.4.

Protein groups identified within the aggregate fractions obtained using the UC-based Brain-Plasma Protocol (BPP) and SEP separately and protein groups "shared" by the two enrichment methods. The shared fraction included only the protein isoforms in common and not the parental proteins. The Coverage% values were normally distributed and the mean was used for comparison.

BPP: Brain-Plasma Protocol

SEP: Seprion PAD-Beads extraction method

PSM: Peptide-Spectrum Match

Total spectra: number of spectra acquired at MS2 level

Total PSMs: number of Peptide-Spectrum Matches identified by PD

ID rate% (Identification rate): PSMs/total spectra

Protein groups: number of proteins identified by the Parsimony Principle in PD

Coverage%/protein group: the coverage% per protein identified (mean)

#Unique Peptides/protein group: number of unique peptides per protein identified (median)

#Peptides/protein group: number of peptides (unique and not unique) per protein identified (median)

#PSMs/protein group: number of PSMs per protein identified (median)

3.5.1.1 Analysis of the BPP and SEP shared proteins

From the 225 proteins identified in both BPP and SEP datasets, PANTHER identified 196 genes accounting for 138 hits within the CC terms (Figure 3.5A). From the 196 genes identified by PANTHER, 47 (34.1% of the total hits) were classified as cell part (i.e. any constituent part of a cell), with the majority being for intracellular proteins (35 genes, 67.3% of the total hits within cell part) and as extracellular region (47 genes, 37.1% of the total hits) (Figure 3.5A). The other two main CC were macromolecular complexes (15.2%) and organelles (10.9%), where ten genes (62.50% of the total hits within organelles) were classified as cytoskeleton (Figure 3.5A). The 196 genes were submitted to Webgestalt for pathway functional analysis on KEGG database and among the top10 enriched pathways (Table 3.5), six were linked to infection and immune response (staphylococcus aureus infection, pertussis, systemic lupus erythematosus, pathogenic Escherichia Coli infection, complement, platelet activation) and the remaining four were related to cell migration (adhesion), extra-cellular matrix and prion disease. As highlighted by PANTHER analysis, the cellular components involved in these pathways were closely related to extracellular region and cell part terms.

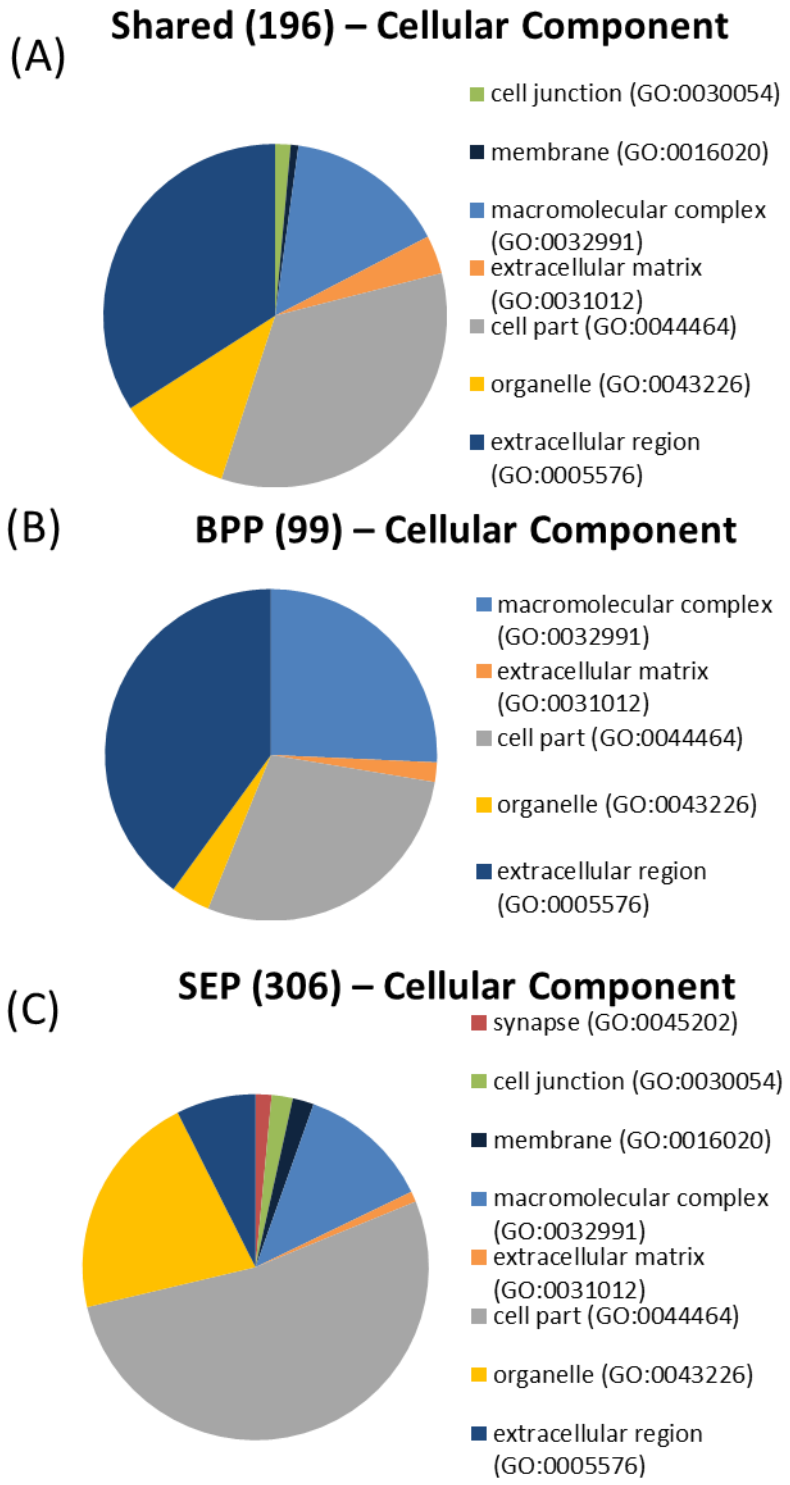


Figure 3.5. Pie charts showing cellular component (CC; PANTHER database) terms identified using BPP, SEP and using both enrichment methodologies (shared).

(A) Terms represented in both BPP and SEP shared proteins groups. (B) Terms represented in the BPP protein groups. (C) Terms represented in the SEP protein groups. The number of genes identified by PANTHER for each group of proteins is reported in bold and under brackets.

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04610	Complement and coagulation cascades	79	40	1.23	32.42	0.00E+00	0.00E+00
hsa05150	*Staphylococcus aureus* infection	56	13	0.87	14.86	1.72E-12	2.60E-10
hsa05020	Prion diseases	35	8	0.55	14.63	4.53E-08	4.58E-06
hsa05133	Pertussis	76	10	1.19	8.42	2.30E-07	1.74E-05
hsa04510	Focal adhesion	203	14	3.17	4.42	2.78E-06	1.69E-04
hsa05322	Systemic lupus erythematosus	135	11	2.11	5.22	7.20E-06	3.64E-04
hsa05130	Pathogenic *Escherichia coli* infection	55	7	0.86	8.15	2.04E-05	8.85E-04
hsa04145	Phagosome	154	11	2.41	4.57	2.52E-05	9.56E-04
hsa04512	ECM-receptor interaction	82	7	1.28	5.47	2.71E-04	9.13E-03
hsa04611	Platelet activation	122	8	1.91	4.20	6.03E-04	1.83E-02

Table 3.5. Top 10 enriched KEGG pathways in the SEP and BPP shared proteins list.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: Number of genes included in the given KEGG pathway in Homo sapiens

Number of genes matched in the category: Number of genes in the submitted list also identified in the given KEGG pathway

Expected number of genes in the category: Calculated as the number of genes submitted divided by the number of the total Homo sapiens genes involved in all the KEGG pathway database and multiplied by the number of genes in the category

Ratio of enrichment: Calculated as the number of genes matched in the category divided by the Expected number of genes in the category

p-Value: Statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): Calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

3.5.1.2 Functional analysis of the proteins data identified using BPP

From the 144 proteins represented in the BPP proteins data, PANTHER identified 99 genes accounting for 105 hits within the CC terms (Figure 3.5B). From these 99 genes, 42 (40% of the total hits) were classified as part of the extracellular region, 30 (28.6%) as cell part and 27 (25.7%) as macromolecular complex, while the rest belonged to extracellular matrix (2, 1.9%) and organelles (4, 3.8%) (Figure 3.5B). 8 genes (26.7%) of the cell part component were intracellular proteins and 22 (73.30%) were plasma membrane proteins, while 1 (20%) of the organelle component was part of the cytoskeleton. The pathway functional analysis of the BPP proteins data (Table 3.6) revealed a significant representation of features involved in endocrine, metabolic and cell-signalling regulation (PPAR, thyroid hormone synthesis, lysine degradation, carbon metabolism, citrate-TCA cycle) and pathways specific to ECM-receptor interaction, phagosome and adipocytokine signalling.

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa00020	Citrate cycle (TCA cycle)	30	3	0.13	23.26	2.69E-04	0.06
hsa04512	ECM-receptor interaction	82	4	0.35	11.35	3.85E-04	0.06
hsa04145	Phagosome	154	4	0.66	6.04	4.00E-03	0.40
hsa01200	Carbon metabolism	114	3	0.49	6.12	1.25E-02	0.95
hsa05144	Malaria	49	2	0.21	9.50	1.85E-02	1.00
hsa05130	Pathogenic Escherichia coli infection	55	2	0.24	8.46	2.30E-02	1.00
hsa00310	Lysine degradation	59	2	0.25	7.89	2.63E-02	1.00
hsa04920	Adipocytokine signaling pathway	70	2	0.30	6.65	3.60E-02	1.00
hsa03320	PPAR signaling pathway	72	2	0.31	6.46	3.79E-02	1.00

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04918	Thyroid hormone synthesis	74	2	0.32	6.29	3.99E-02	1.00

Table 3.6. Top 10 enriched KEGG pathways in the BPP proteins list.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: Number of genes included in the given KEGG pathway in Homo sapiens

Number of genes matched in the category: Number of genes in the submitted list also identified in the given KEGG pathway

Expected number of genes in the category: Calculated as the number of genes submitted divided by the number of the total Homo sapiens genes involved in all the KEGG pathway database and multiplied by the number of genes in the category

Ratio of enrichment: Calculated as the number of genes matched in the category divided by the Expected number of genes in the category

p-Value: Statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): Calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

3.5.1.3 Functional analysis of the protein data identified using BPP

From the 380 proteins represented in the SEP proteins dataset, PANTHER identified 306 genes accounting for 202 hits within the CC (Figure 3.5C). From the 202 genes, the main component was cell part with 106 genes (52.5% of the total hits), of which 102 (79.7%) were intracellular and 23 (18%) were plasma membrane proteins, followed by organelle (43 genes, 21.3% of total hits) and macromolecular complex (25, 12.4%) components, while extracellular region, cell junction, membrane, synapse and extracellular matrix accounted for 28 genes (13.9% of the total hits) (Figure 3.5C). Most of the organelle sub-component proteins were related to the cytoskeleton (25, 53.2%) (Figure 3.5C). The functional analysis of the SEP unique proteins data (Table 3.7) identified features related to immune activation (chemokine signaling pathway, focal adhesion, shigellosis, platelet activation) and to cell function and structure (tight junction, regulation of actin cytoskeleton, vascular smooth muscle contraction). All pathways showed enrichments with a FDR < 0.1.

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04810	Regulation of actin cytoskeleton	216	27	5.23	5.16	1.11E-12	3.37E-10
hsa04611	Platelet activation	122	16	2.95	5.42	3.05E-08	3.67E-06
hsa05131	Shigellosis	65	12	1.57	7.62	3.64E-08	3.67E-06
hsa04270	Vascular smooth muscle contraction	121	15	2.93	5.12	1.80E-07	1.24E-05
hsa04921	Oxytocin signaling pathway	159	17	3.85	4.42	2.36E-07	1.24E-05
hsa04144	Endocytosis	260	22	6.30	3.49	2.46E-07	1.24E-05
hsa05130	Pathogenic Escherichia coli infection	55	10	1.33	7.51	6.07E-07	2.63E-05
hsa04530	Tight junction	139	14	3.37	4.16	5.87E-06	2.23E-04
hsa04510	Focal adhesion	203	17	4.92	3.46	7.27E-06	2.45E-04

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04062	Chemokine signaling pathway	187	16	4.53	3.53	1.04E-05	3.16E-04

Table 3.7. Top 10 enriched KEGG pathways in the SEP unique proteins list.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: Number of genes included in the given KEGG pathway in Homo sapiens

Number of genes matched in the category: Number of genes in the submitted list also identified in the given KEGG pathway

Expected number of genes in the category: Calculated as the number of genes submitted divided by the number of the total Homo sapiens genes involved in all the KEGG pathway database and multiplied by the number of genes in the category

Ratio of enrichment: Calculated as the number of genes matched in the category divided by the Expected number of genes in the category

p-Value: Statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): Calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

3.5.2 Immunogold-Transmission Electron Microscopy (IG-TEM)

It has been reported that the aggregation process leading to the formation of macromolecular structures within a fluid can be visualized by transmission electron microscopy (TEM), as stated by Ross and Poirier (55). In line with this observation, Immunogold Transmission Electron Microscopy (IG-TEM) was employed for the purpose of confirming the presence of macromolecular formations in the fractions extracted from HC-PPS using the ultracentrifugation (BPP) and Seprion (SEP) enrichment protocols and to establish the presence of Nf within these macromolecular structures. Aggregate fractions for TEM analysis were obtained using SEP (Paragraph 2.2.1) and using BPP (Paragraph 2.2.2.1). The use of TEM was intended to demonstrate the presence of electron-dense (and likely heterogeneous) macrostructures representing the high molecular weight aggregates identified using gel-based methods after enrichment of HC-PPS. Using TEM it was not possible to infer any information on the overall composition of these macromolecular formations apart from the presence of Nf proteins based on the binding of gold particles. As a first step, optimal conditions for TEM analysis including incubation time for sample absorption onto the grid, negative stain using Uranyl Acetate or Ammonium Molybdate and antibody dilutions were tested (Figure 3.6).

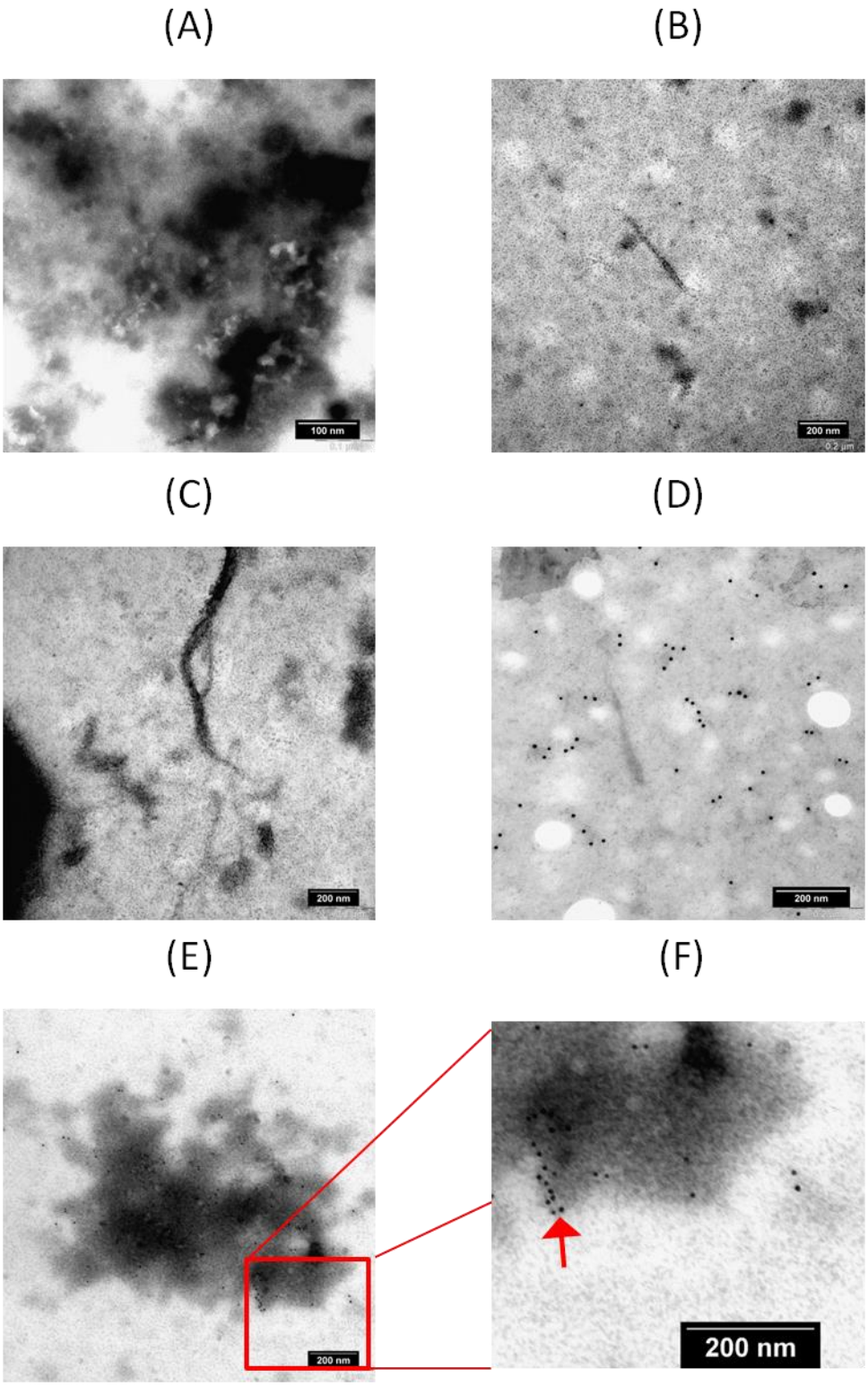


Figure 3.6. Optimisation of the experimental conditions for the study of NCHs using Transmission Electron Microscopy (TEM) and Immunogold-TEM (IG-TEM).

(A) Negative staining: BPP-enriched fraction in the glow discharged grid incubated for 30 seconds in Ammonium Molybdate (AM). (B) Negative staining: BPP-enriched fraction in the glow discharged grid incubated for 10 seconds in AM and immunolabelling with anti-NfM diluted 1:100. (C) Negative staining: BPP enriched fraction incubated for 10 seconds in AM and immunolabelling

with anti-NfL diluted 1:100. (D) Negative staining: bovine NfH (bNfH) incubated for 10 seconds in Uranyl Acetate (UA) and immunolabelling with anti-NfH diluted 1:100. (E) Negative staining: BPP-enriched fraction incubated for 10 seconds in UA with anti-NfH diluted 1:200. (F) Magnification of a small part of E showing gold particles binding to the electron dense structure. Scale bars are indicated at the low right corner of each micrograph.

In the negative staining with 30 seconds incubation (Figure 3.6A), the electron-dense material appeared heterogeneous and diffused on to the whole grid possibly also as a result of over-loading. The incubation time was decreased to 10 seconds (Figure 3.6B). 10 seconds incubation time with Ammonium Molybdate (AM) gave a good contrast but anti-NfM and anti-NfL gold particles linked to the secondary antibodies were not detectable (Figure 3.6B, C). Negative staining with Uranyl Acetate (UA) showed a lighter contrast, with immune-gold particles detection of bovine NfH (bNfH) and of highly electron-dense macrostructures (Figure 3.6D, E, F). For primary antibodies dilutions, anti-NfL, anti-NfM and anti-NfH antibodies were initially tested at 1:100 to avoid a high background. We could not detect any NfL and NfM signal, while NfH in both bNfH and electrondense aggregates was detected at both 1:100 and 1:200 dilutions (1:200 was set as final dilution for the lowest background).

IG-TEM of the BPP fraction showed the presence of globular amorphous formations (Figure 3.7A) and all Nf isoforms were detected (Figure 3.7A, C and E). While NfM (Figure 3.7C) and NfL (Figure 3.7E) immune-gold signals were primarily represented outside the aggregates, NfH (Figure 3.7A) was mainly detected within electron-dense formations, which is in line with the consistent detection of anti-NfH in high molecular weight bands detected in our WB experiments (Figure 3.3A).

The grids coated with SEP-enriched fractions showed electron-dense macromolecular formations confirming the presence of protein aggregates in the SEP-processed sample (Figure 3.7B, D and F). However, while BPP-enriched fractions appeared with more globular electron-dense features (Figure 3.7A) the final product of the SEP enrichment appeared more as a dense network of fibers, with evidence of salt retention (Figure 3.7D). The appearance of a fiber network could be due to the lack of any pre-treatment in the SEP enrichment method, while the presence of salts to the specifics of the elution buffer (0.1 M NaOH,

0.1% Triton X-100 and Tris-HCl). Both NfH (Figure 3.7B) and NfM (Figure 3.7D) gold particles were within electron-dense formations but also outside which is in line with the high molecular weight (MW) bands detected by WB (Figure 3.2B). A few NfL (Figure 3.7F) gold particles in contrast with WB data where no NfL band was detected (Figure 3.2C), were detected outside the electron-dense areas.

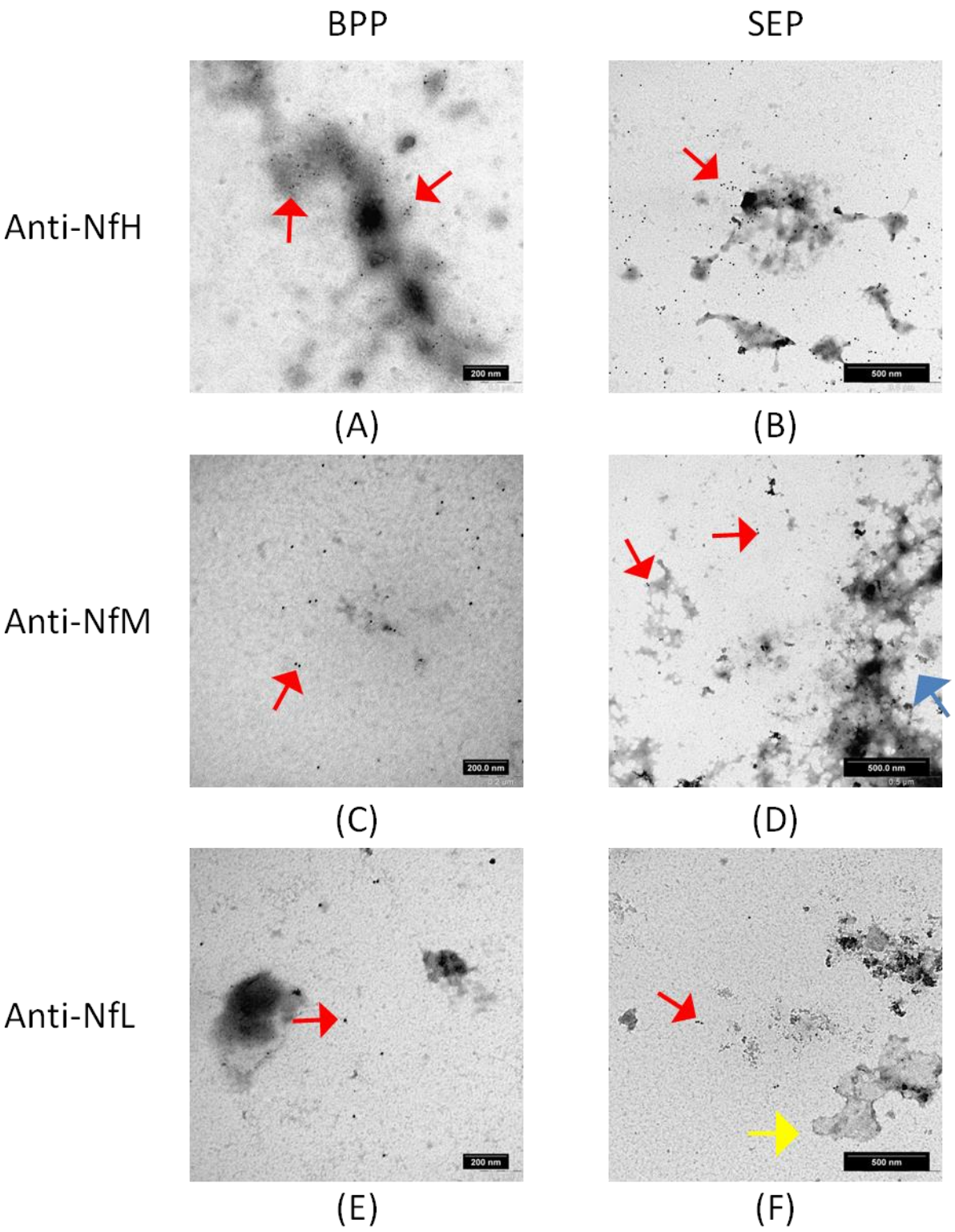


Figure 3.7. IG-TEM micrographs to test anti-Nf proteins antibodies on protein aggregates from HC-PPS extracted using BPP and SEP.

The micrographs displayed the presence of amorphous aggregates for both enrichment methods with the gold particles (red arrows) showing detection of NfH (A and B), NfM (C and D) and NfL (E and F). SEP sample showed also the presence of a fibrous network (blue arrow) (D) and impurities (yellow arrow) (F). Scale bars are indicated at the low right corner of each micrograph.

3.5.3 Plasma aggregates extraction: general considerations

To investigate the composition and functional properties of circulating Neurofilament-Containing Heteroaggregates (NCHs), understanding how selective the chosen enrichment methods are in the process of isolation of these formations is of fundamental importance. Sedimentation and binders exploit different physical properties in the enrichment process, with the SEP protocol based on affinity for proteins involved in different levels of aggregation (110). SEP enrichment allowed the identification of a higher number of proteins, an observation that may be explained by the ligand's broader affinity to any conformational change, which affect protein-protein interactions. This method may in fact be more suitable to the isolation of disordered monomers and/or low order multimers and not only of large MW heteroaggregates. On the other hand, density was the main feature exploited in the enrichment process by BPP, where large and dense aggregates are expected to sediment with ultracentrifugation (UC) while low order multimers (which may function as seeds of larger protein aggregates) may escape detection. Overall BPP separation appeared to be more stringent, generating a smaller and possibly less complex proteome, while acquiring, on average, more spectra for the same peptide. This means that although the BPP method produced less information from the proteomics analysis, identifying a lower number of protein groups, it likely generated less false positive information in protein identification compared to SEP.

Qualitative differences in the protein pools within aggregates extracted using UC and SEP have emerged. MS proteomics showed that aggregates extracted with both methodologies mostly harbour cellular components, although interspersed with extracellular matrix proteins. Functionally, both extraction techniques identified proteins networked within molecular cascades crucial to inflammatory processes, particularly complement factors, along with phagosome and prion-related proteins. UC aggregates are specifically enriched with proteins involved in endocrine, metabolic and cell-signaling regulation. It could be speculated that these hetero-complexes naturally express biological features which are essential in maintaining homeostasis, like inflammation and metabolism. It can also be speculated that circulating protein assemblies may act as sinks for the sequestration of molecular factors that may have otherwise detrimental effects.

Despite the different biological information that could be obtained from the analysis of the protein pools obtained using the two methods, BPP was more reliable for protein identification by LC-MS/MS and showed a higher compatibility to orthogonal techniques. These observations were crucial in the decision of adopting BPP as the method choice to carry out other aspects of this study.

With regards to the immunogold profiles of Nf isoforms, both SEP and BPP extraction method showed a predominant co-localization of NfH and aggregates, a pattern that was also reproduced in NfM micrograph (Figure 3.7C and D), while NfL gold particles appeared only outside the electrondense structures (Figure 3.7E and F). It is possible that the different distribution of Nf isoforms within and outside protein aggregates may be explained by the interactions between Nf and other proteins, which depend on conformation and on Post-Translational Modifications (PTMs), and structural properties which can also affect binding with the Seprion ligand (111). For the purpose of the research work detailed in this Thesis and to visualise the presence of macromolecular structures in the fractions obtained by SEP and BPP, we have undertaken IG-TEM optimization process with negative staining, but the lack of quality controls does not allow a better interpretation of the amorphous and fibrous electron-dense features identified in our micrographs.

3.6 Effect of urea on plasma protein aggregates from HC and ALS patients

Lu et al. have reported that urea has a disruptive effect on protein aggregates and the use of this denaturing agent increases NfH detection by solubilisation of NfH-containing aggregates which in turn reduces NfH epitope masking in protein assemblies; this process is likely to increase the overall sensitivity of enzyme-linked immune assays (ELISA) for NFH detection (53). In line with this protocol, we have used 8 M urea to resuspend the final pellet obtained by BPP enrichment from HC-PPS and ALS-PPS (Figure 3.8). ALS-PPS was created pooling plasma samples from six ALS patients with heterogeneous clinical features, including fast and slow disease progression rate (Table 3.8).

Sample ID	Gender	Ethnicity	Age at sampling (year)	NfH (ng/ml)	El-Escorial	Clinical onset	ALSFRS-R	Progression Rate at visit
ALS1-PPS	M	Caucasian	76.77	NA	Definite ALS	R/B	37	5.979
ALS2-PPS	M	Caucasian	59.41 – 59.56	NA	Definite ALS	R	25 – 39	1.839 – 3.665
ALS3-PPS	F	Caucasian	71.56	NA	Probable ALS	R\B	33	3.106
ALS4-PPS	M	Caucasian	60.67	NA	Suspected ALS	LL	NA	0.000
ALS5-PPS	M	Afro-Caribbean	46.12	40.6	Possible ALS	LL	47	0.004
ALS6-PPS	M	Caucasian	78.48	NA	Possible ALS	B	46	0.004

Table 3.8. ALS-Pooled Plasma Sample (ALS-PPS): clinical features and demographics of the selected ALS patients.

For ALS2, we have identified samples from more than one time point (age at sampling time points are reported)

Sample ID: Identification code

Gender: M= Male; F= Female

Age at sampling: age of the patient at the time of blood sampling (age-range shown if more than one time point sample was used)

NfH (ng/ml): level of NfH expressed in ng/ml; NA= data not available

El-Escorial: diagnostic classification of ALS (88)

Clinical Onset (site of initial clinical signs): R= Respiratory; B= Bulbar; LL= Lower Limbs

ALS Functional Rating Scale revised (ALSFRS-R): level of neurological impairment across different clinical domains (1-48, higher neurological impairment with lower values); NA= data not available

Rate of progression at sampling time point: calculated as 48 - ALSFRS-R score at sampling time/disease duration from onset of symptoms to sampling time.

WB for NfH detection showed high MW bands for both ALS-PPS and HC-PPS enriched fractions, confirming the presence of NCHs in both healthy and ALS pooled samples (Figure 3.8). The two pooled samples displayed the same WB NfH profile, with the ALS-PPS showing a higher NfH intensity in all conditions tested. The NfH bands detected in the ALS-PPS fraction showed different intensities depending on the buffer used for solubilisation (e.g. PBS vs urea). 8 M urea increased of approximately 2-fold the intensities of the 238, 117 and 55 KDa bands, while reducing the intensity of the high MW band situated above the expected NfH MW (238 KDa; Figure 3.8). A similar trend was present in the HC-PPS fraction resuspended in 8 M urea (Figure 3.8). This observation may indicate that NCHs composition differs in ALS compared to healthy control, hence their susceptibility to urea. Extension of urea treatment to a larger number of samples from healthy controls and ALS patients will be needed to confirm this preliminary observation. We have therefore used 8 M urea in experiment of subilisation of BPP final pellets.

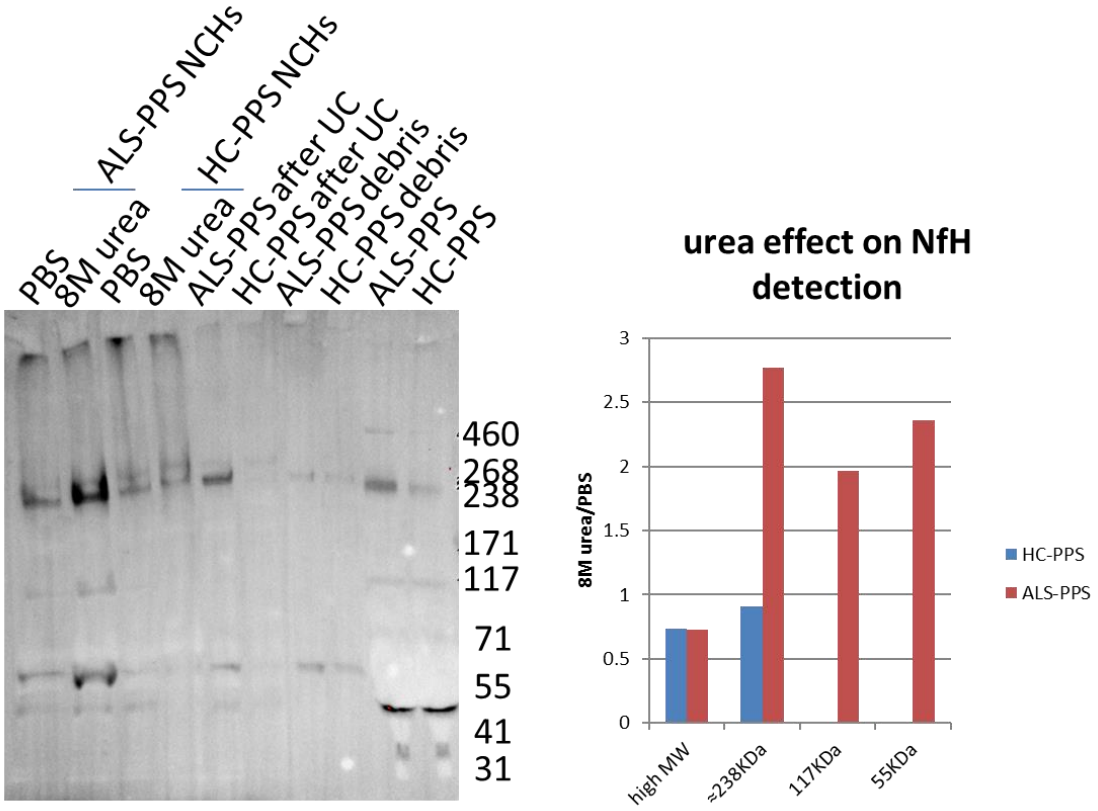


Figure 3.8. Western blot (WB) for NfH detection: urea effect on protein aggregates after BPP enrichment.

WB shows NfH content after resuspension in PBS and 8M urea of the BPP-enriched fractions obtained from Healthy Conrol-Pooled Plasma Sample (HC-PPS) and ALS-Pooled Plasma Sample (ALS-PPS). ALS-PPS enriched fractions are represented in lane 1 and 2 (from left to right), using 8M urea (1) and PBS (2) to resupend the sample after enrichment process. HC-PPS enriched fractions are represented in lane 3 and 4, using 8M urea (3) and PBS (4) as buffer. Lanes 5 and 6 show NfH content, in ALS-PPS (5) and HC-PPS (6) samples, after the first round of ultracentrifugation (UC) in the BPP protocol. Lanes 7 and 8 show the NfH content in the pellet after the benchtop centrifugation step, for ALS-PPS and HC-PPS respectively, while 9 and 10 show the NfH content in HC-PPS and ALS-PPS (before enrichment). The effect of the urea buffer in NfH detection is showed in the chart, where the ratio between the band intensities obtained using urea and PBS is plotted for both HC-PPS and ALS-PPS samples.

3.7 An alternative method for NCHs enrichment by ultracentrifugation: Brain Only Protocol (BOP)

The Brain Only Protocol (BOP) protocol was used to extract protein aggregates from brain for the TEM applications. The protocol was based on previously published data (Greenberg and Davied paper (97)) and adapted to NCHs extraction alongside the aforementioned BPP protocol.

3.7.1 NCHs extraction from brain using BPP and BOP

Brain1 and Brain2 samples were homogenised and mixed as described in Paragraph 2.1.2.1 BOP and BPP protocols (Paragraph 2.2.2.1 and 2.2.2.2) were then used for NCHs extraction from the brain homogenate. Following extraction, quantification of total protein concentration showed a lower protein concentration for the fraction enriched in protein aggregates obtained by BOP ((BPP= 0.336 µg/µl; BOP= 0.297 µg/µl). This difference was probably related to the presence of a second UC step with a sucrose cushion in the BOP protocol. To evaluate further the differences between BPP and BOP, SDS-PAGE and WB against NfH were performed loading 5 µg of the final extraction products (NCHs enriched), together with the original brain sample homogenate and the intermediate products of the two extraction processes (brain debris collected after the benchtop centrifugation, supernatant and pellet after the first UC step in the two protocols; Figure 3.9). The SDS-PAGE stained with Imperial Blue Comassie showed very faint bands, limiting any further in-depth evaluation of the enrichment processes (Figure 3.9). Conversely, it was possible to evaluate the presence of NfH in the BOP and BPP NCHs enriched fractions, with the BPP

NCHs enriched fraction showing a higher intensity of NfH staining compared to the BOP NCHs enriched fraction (Figure 3.9, lanes 1 and 4).

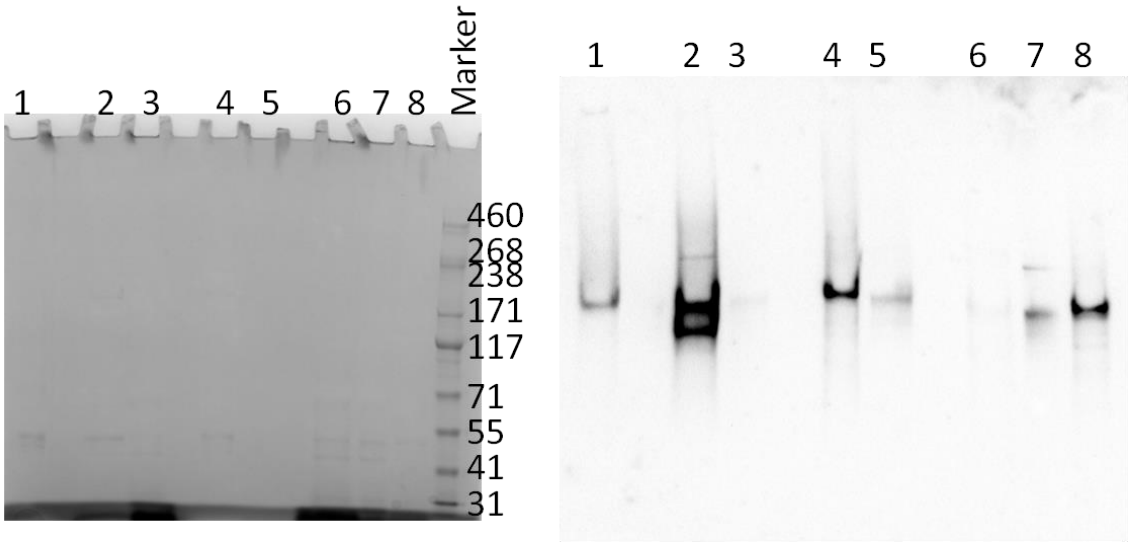


Figure 3.9. Analysis of brain NCHs extracted using BOP and BPP: SDS-PAGE and anti-NfH WB.

NCHs enriched fractions isolated using BOP and BPP protocols, together with samples collected from the intermediate steps of the extraction oricesses were loaded (5 µg) onto a 3-8% Tris-Acetate gels (SDS-PAGE, left hand side) and used for Western Blot to test NfH expression (right-hand side). Few and faint bands were detected at MW lower than 55 KDa on the SDS-PAGE (left-side), while NfH was detected at different MW bands in the WB (right-hand side). BPP NCHs enriched fraction (lane 4) showed a more intense NfH band than BOP NCHs enriched fraction (lane 1). Lanes: 1, BOP NCHs enriched fraction; 2, BOP pellet after the first ultracentrifugation (UC) step; 3, BOP supernatant after the first UC step; 4, BPP NCHs enriched fraction; 5, BPP pellet after the first UC step; 6, BPP supernatant after the first UC step; 7, Brain sample; 8, Brain sample debris; Marker, HiMark MW marker.

After testing the BPP and BOP extraction methods on brain samples, BOP was also used for NCHs extraction from HC-PPS and ALS-PPS (Figure 3.10) and compared to the NCHs extraction obtained with BPP (Figure 3.8). NfH was detected in un-processed pooled plasma (Figure 3.10 lanes 9 and 10). High intensity NfH bands were detected in plasma debris, the pellet obtained after the initial plasma benchtop centrifugation at 21000xg (Figure 3.10, lanes 7 and 8), (Figure 3.10, lanes 9 and 10). In the final extraction products by BOP, NfH was detected in the NCHs enriched fractions at a single 238/268 KDa band and also at higher MW (Figure 3.10, lanes 1 and 2), while bands lower than 238 KDa were not present as previously shown in the final products of extraction by BPP (Figure

3.8, lanes 1 and 2). In the supernatant and the pellet collected after the first UC step, only bands below 71 KDa were detected by the anti-NfH antibody (Figure 3.10, lanes 3-6). Stringency in the BOP extraction method appears to increase with the second UC step (Figure 3.10, lanes 1 and 2), where NfH bands at the expected and at much higher molecular weight were detected (and not the two bands at 55 and 41 KDa present after the first UC step).

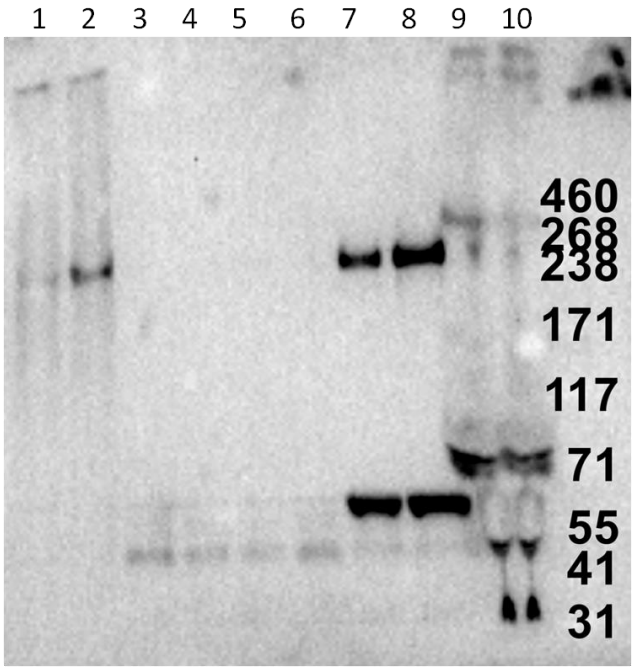


Figure 3.10. NfH presence in plasma NCHs enriched fractions from ALS-PPS and HC-PPS isolated by BOP.

Western Blot (WB) analysis shows NfH at the expected MW (238/268 KDa) in both HC-PPS and ALS-PPS (lanes 9 and 10), in the plasma debris (lanes 7 and 8) and in the final products (lanes 1 and 2) of extraction by BOP. NfH was also detected at a lower MW (< 71 KDa) in lanes 3 to 10, but not in the final NCHs enriched fractions (lanes 1 and 2) obtained by BOP. Only the final products of extraction by BOP (lane 1 and 2) showed high MW bands likely to represent NCHs. Lanes: 1, HC-PPS NCHs enriched fraction; 2, ALS-PPS NCHs enriched fraction; 3, HC-PPS supernatant after the second ultracentrifugation (UC); 4, ALS-PPS supernatant after the second UC; 5, HC-PPS pellet after the first UC; 6, ALS-PPS pellet after the second UC; 7, HC-PPS plasma pellet after benchtop centrifugation (debris); 8, ALS-PPS plasma debris; 9, HC-PPS; 10, ALS-PPS.

After WB evaluation of the NCHs extraction protocols and having confirmed the presence of NCHs in the final products of extraction from both plasma and brain using BOP, LC-MS/MS proteomics was performed to compare the composition of NCHs obtained using BPP and BOP enrichment from brain and plasma. The

final extraction products from HC-PPS and ALS-PPS were loaded into a gradient gel and after electrophoresis, ten fractions were cut from each lane and in-gel trypsin digestion performed. For brain NCHs proteomic analysis, BOP and BPP NCHs enriched fractions were loaded into the gel and electrophoresis blocked to concentrate the samples in one single band after they had entered completely in the upper edge of the gel. The two bands, one containing BOP NCHs and the other BPP NCHs, were cut and in-gel trypsin digestion was performed. Following protein identification as previously described, the protein content of NCHs obtained using BOP and BPP protocols were compared for each sample type (ALS-PSS, HC-PSS and Brain) (Figure 3.11, Table 3.9 and Table 3.10).

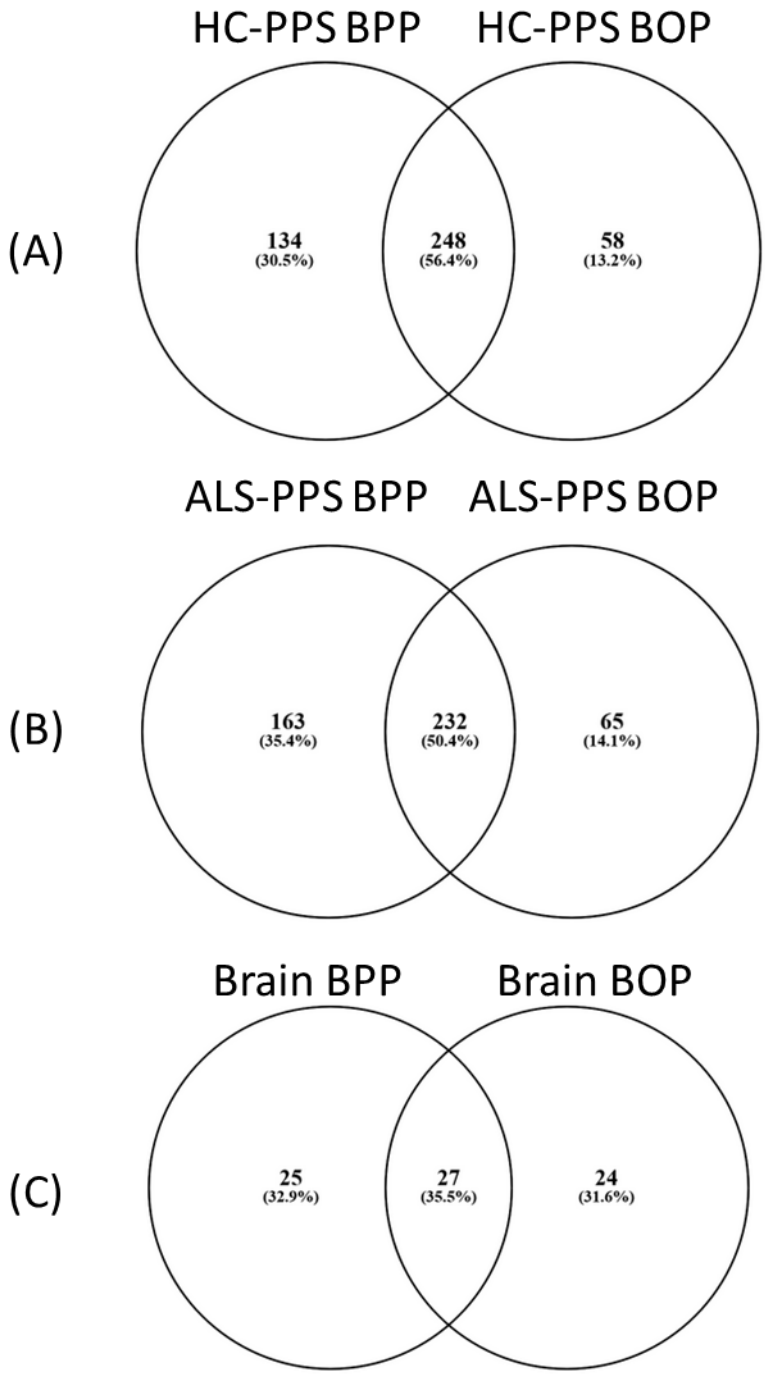


Figure 3.11. Number of proteins identified in NCHs extracted from ALS-PSS, HC-PSS and Brain and their distribution based on BPP and BOP enrichment.

The Venn diagrams show the number of NCHs proteins identified using BPP and BOP extraction methods for (A) HC-PPS, (B) ALS-PPS and (C) Brain sample (mixed homogenate of Brain1 and Brain2). Within each section of the Venn diagrams, we report the number of proteins (and their percentage of the total proteins detected under brackets) that are either unique to one of the extraction methods or shared by the two (only proteins identified by a minimum of 2 peptides and a peptide count only in the top scored proteins). HC-PPS BPP: NCHs enriched fraction obtained from HC-PPS sample with BPP protocol; HC-PPS BOP: NCHs enriched fraction from HC-PPS

sample with BOP protocol; ALS-PPS BPP: NCHs enriched fraction from ALS-PPS sample with BPP protocol; ALS-PPS BOP: NCHs enriched fraction from ALS-PPS sample with BOP protocol; Brain BPP: NCHs enriched fraction from Brain sample with BPP protocol; Brain BOP: NCHs enriched fraction from Brain sample with BOP protocol.

Protein identification in HC pooled samples detected 382 proteins by BPP and 306 by BOP, with 248 proteins in common, which corresponded to the 56.4% of the total amount of identified proteins using both extraction methods (Figure 3.11A). The enriched fraction from ALS-PPS showed the same trend, with 395 proteins detected after BPP enrichment and 297 after BOP, and 232 (50.4% of the total proteins identified) proteins in common (Figure 3.11B). Brain NCHs showed a distinctively different picture, with 52 and 51 proteins identified in the BPP and BOP-NCHs enriched fractions respectively, and only 27 (35.5%) proteins in common (Figure 3.11C). A more informative way to look at LC-MS/MS performance in the analysis of the fractions produced using BPP and BOP enrichment methods to look broadly at the main *protein identification features* of each sub-group showed in Figure 3.11 (Table 3.9 and 10). In HC-PPS and ALS-PPS plasma samples (Table 3.9), BPP had slightly better outcomes (higher ID rate and coverage). However, when considering the fraction of shared proteins, BPP outperformed BOP showing a considerable higher coverage, number of unique peptides, total peptides and PSMs per protein group, in both ALS and HC samples. With regards to the data from brain NCHs (Table 3.10), BOP showed higher coverage per protein group, while BPP performed better in number of unique peptides, total peptides, PSMs and identification score per protein group in the shared proteins fraction. BPP showed an ID rate% that was almost double that of BOP (27.87% in BPP against 15.17% in BOP) which indicated that it was possible to obtain more PSMs, and therefore more information, from the same amount of spectra acquired by the Mass Spectrometer (possibly because of a higher quality of the spectra generated).

While LC-MS/MS analysis provided an in-depth profile of the protein composition of the enriched fractions obtained using different extraction protocols, it did not identify Nf proteins in the BPP and BOP- plasma NCHs, in contrast to the confirmation of Nf expression obtained by WB in these samples. Conversely, the analysis of brain samples enriched fractions obtained with both extraction protocols showed all Nf proteins (NfH, NfM and NfL). NfH was detected as NfH

isoform 2, with a 17.96% total sequence coverage by BPP (a score of 179) and a 12.67% total sequence coverage by BOP (a score of 53.7). NfM was detected by BPP with a 19.76% coverage (a score of 206.02) and by BOP with 14.74% coverage (a score of 66.55). NfL showed the highest difference of the coverage detected by the two extraction methods, with a 24.49% coverage and a score of 64.41 for BPP and a 5.16% coverage and a score of 9.70 for BOP. The discrepancy in the LC-MS/MS analysis of Nf in brain and in plasma samples may be explained by the differences in abundance of these proteins in brain and plasma but also by the way proteomics identifies and process signals related to defined proteins. The higher Nf abundance in brain enriched fraction compared to plasma, clearly translate into a higher level of detection Nf peptides in the ionic form (parental ion) during the first MS scan (MS1), which is then further enhanced by peptide fragmentation and spectrum acquisition in the second MS scan (MS2). In consideration of the better performance of BPP as shown by the overall protein coverage, number of unique peptides, total peptides, PSMs and score for protein group (Table 3.9 and Table 3.10), we have selected this enrichment protocol for those experiments where quantitative proteomics is used for further characterization of NCHs in brain and plasma. To achieve this, as reported in the Chapter 5, we will use a pioneering proteomic technique, which allows simultaneous analysis of brain tissue and plasma samples, quantitative determination of protein expression and the identification of low-abundance proteins in the plasma matrix that are known to be expressed in and be relevant in brain function.

Sample	Total spectra	Total PSMs	ID rate%	Protein groups	Coverage%/ protein group (mean)	#Unique Peptides/ protein group (median)	#Peptides/ protein group (median)	#PSMs/ protein group (median)
HC-PPS BPP	325811	24577	7.54	382	29.09	5.00	7.00	10.00
HC-PPS BOP	296876	20834	7.02	306	28.67	5.00	6.00	13.00
HC-PPS BPP (shared)				248	35.69	8.00	9.00	20.50
HC-PPS BOP (shared)				248	31.19	6.00	8.00	18.00
ALS-PPS BPP	348125	32808	9.42	395	30.33	5.00	6.00	12.00
ALS-PPS BOP	283357	20469	7.22	297	28.69	5.00	6.00	13.00

Sample	Total spectra	Total PSMs	ID rate%	Protein groups	Coverage%/ protein group (mean)	#Unique Peptides/ protein group (median)	#Peptides/ protein group (median)	#PSMs/ protein group (median)
ALS-PPS BPP (shared)				232	38.16	9.00	11.00	28.00
ALS-PPS BOP (shared)				232	31.38	7.00	8.00	20.00

Table 3.9. Liquid Chromatography coupled with tandem Mass Spectrometry (LC-MS/MS) analysis of the NCHs enriched fractions obtained from the Healthy Control-Pooled Plasma Sample (HC-PPS) and ALS-Pooled Plasma Sample (ALS-PPS) using the BPP and BOP enrichment protocols: proteomic performance and protein groups defined by Proteome Discoverer 1.4 (PD).

Protein groups (proteins in which the peptides identified by LC-MS/MS are grouped as result of sequence redundancy) identified within the aggregate fractions obtained from HC-PPS and ALS-PPS using BPP and BOP methods. The shared fractions included only the protein isoforms in common and not the parental proteins. The Coverage% values were normally distributed and the mean was used for comparison.

HC-PPS BPP: NCHs enriched fraction obtained from HC-PPS by BPP protocol

HC-PPS BOP: NCHs enriched fraction obtained from HC-PPS by BOP protocol

ALS-PPS BPP: NCHs enriched fraction obtained from ALS-PPS by BPP protocol

ALS-PPS BOP: NCHs enriched fraction obtained from ALS-PPS by BOP protocol

PSMs: Peptide-Spectrum Matches

Total spectra: number of spectra acquired at MS2 level

Total PSMs: number of Peptide-Spectrum Matches identified by PD

ID rate% (Identification rate): PSMs/total spectra

Protein groups: number of proteins identified by the Parsimony Principle in PD

Coverage%/protein group: the coverage% per protein identified (mean)

#Unique Peptides/protein group: number of unique peptides per protein identified (median)

#Peptides/protein group: number of peptides (unique and not unique) per protein identified (median)

#PSMs/protein group: number of PSMs per protein identified (median)

Sample	Total spectra	Total PSMs	ID rate%	Protein groups	Coverage%/ protein group (mean)	#Unique Peptides/ protein group (median)	#Peptides/ protein group (median)	#PSMs/ protein group (median)	Score (mean)
Brain BPP	4916	1370	27.87	52	11.67	5.00	8.00	13.00	59.85
Brain BOP	4939	749	15.17	51	12.81	2.00	6.00	10.00	54.77
Brain BPP (shared)				27	13.91	8.00	10.00	22.00	71.76
Brain BOP (shared)				27	9.92	5.00	7.00	9.00	32.39

Table 3.10. Liquid Chromatography coupled with tandem Mass Spectrometry (LC-MS/MS) analysis of the NCHs enriched fraction obtained from the Brain sample with BPP and BOP enrichment protocols: proteomic performance and protein groups defined by Proteome Discoverer 1.4 (PD).

Protein groups (proteins in which the peptides identified by LC-MS/MS are grouped as result of sequence redundancy) identified within the aggregate fractions obtained from the Brain sample using BPP and BOP methods. The shared fractions included only the protein isoforms in common and not the parental proteins. The Coverage% values were normally distributed and the mean was used for comparison.

Brain BPP: NCHs enriched fraction obtained from Brain sample by BPP protocol

Brain BOP: NCHs enriched fraction obtained from Brain sample by BOP protocol

PSMs: Peptide-Spectrum Matches

Total spectra: number of spectra acquired at MS2 level

Total PSMs: number of Peptide-Spectrum Matches identified by PD

ID rate% (Identification rate): PSMs/total spectra

Protein groups: number of proteins identified by the Parsimony Principle in PD

Coverage%/protein group: the coverage% per protein identified (mean)

#Unique Peptides/protein group: number of unique peptides per protein identified (median)

#Peptides/protein group: number of peptides (unique and not unique) per protein identified (median)

#PSMs/protein group: number of PSMs per protein identified (median)

Score: score given by the SEQUEST-HT algorithm in PD per protein identified (mean)

3.8 Conclusion

In this chapter, we have tested methodologies for the extraction and for the molecular characterization of protein aggregates containing neurofilaments in blood and in brain tissue. To this end, we have used gel-separation and immunodetection by western blot (WB) to confirm the presence of high-molecular weight circulating neurofilament-containing heteroaggregates (NCHs) in blood and applied TEM to obtain a visual confirmation of the presence of these macromolecular structures and of their content in neurofilaments in enriched fractions from pooled plasma. Finally, LC-MS/MS was used to define the protein composition of these aggregates and to look at how these protein profiles differ between healthy and pathological state, as well as between blood and brain tissue.

As the conceptual framework of disease biomarkers evolves, circulating protein aggregates may become increasingly important as they convey information on proteinopathies and may relate to specific organ pathologies. Most of the experimental work produced so far on the detection and quantification of protein aggregates was undertaken in tissues and cell systems (112–115). In the field of neurodegeneration, for example, protein aggregates are formed starting by specific “seed” proteins like Tau, Abeta and amyloid-like polyglutammine which are normally isolated using ultracentrifugation, membrane filter assays and magnetic-bead immunoaffinity pulldowns, with little attention for the wide range of other co-aggregated proteins (112–115). The emerging evidence that different stages of protein aggregation can be seen also in biofluids and the idea that this phenomenon could somehow reproduce brain pathology moved this field of investigation towards the characterisation of the fibrillary and aggregated state of proteins as potential disease biomarkers (116). The development of sound methodologies for the isolation and characterization of protein aggregates from biofluids is therefore critical not only for diagnostic purposes but also for the study of protein aggregates forming as by-products of industrial production protein-based therapies, a major problem in the commercialisation of new monoclonal-antibodies (Ab) therapies (117,118).

Our data not only show unequivocally that Nf proteins are entangled within high MW complexes in plasma as previously anticipated (46,53), but also strengthen

the hypothesis that NfH is more likely to be sequestered into these circulating aggregates, while smaller and perhaps less complex Nf isoforms like NfL do not, in line with the reported NfL linearity when tested by immunodetection in dilution curves compared to the hook effect seen with NfH in the same experiments (47). However, despite the fact that the gel-based separation and WB analysis of Nf isoforms confirm Nf proteins inclusion (particularly of NfH) in high molecular weight formations extracted from plasma, our proteomic MS study did not identify Nf proteins in the NCHs protein mix obtained from plasma, while these three Nf isoforms were detected in brain derived NCHs. A part from the low abundance of Nf proteins in plasma in respect of the entire plasma proteome, failure to identify Nf in NCHs by proteomics may be also explained by the fact that these proteins are subjected to a variety of PTM (119,120). The KSP (Lysine-Serine-Proline) repeats positioned in the Nf tail is the site of most post-translational modifications in Nf peptides which may escape detection using MS-based proteomics (121–123). To test this hypothesis, in a second run of bioinformatic analysis the unmatched spectra were searched against a database inclusive only of Nf sequences, targeting the two main PTMs: phosphorylation and N-glycosilation. This approach did not generate good PSMs as statistical analysis was weakened by the relatively small size of the source database. In fact, due to the redundant nature of the Nf-enriched database, decoy sequences deriving from this database would be biased, estimating a wrong False Discovery Rate (FDR). Another obstacle to the Nf identification by proteomics was the effect of tryptic digestion on KSP repeat. This is known to generate a considerable number of small peptides (between 3 and 5 aminoacids) which are not readily detectable by MS proteomics, in particular in the case of NfH where KSP repeats are more abundant than in other Nf isoforms. It is also possible that plasma Nf escaped detection by proteomics because of lower digestion yield by trypsin, generating peptides that are too big for the parameters set for proteomics analysis (69). To this end, it can be hypothesized that Nf found in circulating protein complexes may have a modified protease resistance which conditions their propensity to form aggregates. This concept will be further explored in the work detailed in this Thesis.

In this part of the study, the project was focused on assessing the existence of blood NCHs, in line with the observation that protein aggregation and the

formation of NCHs may occur physiologically. It is possible to hypothesise that any organ-specific or systemic pathological state involving protein catabolism or homeostasis may be reflected on the composition and eventually conformation of these circulating complexes, providing a potential disease signature for clinical monitoring and therapeutics. Also, in addition to gel-based separation which has allowed the sizing of these circulating complexes, TEM has provided further insight into the structure of these macromolecular formations whose appearance on EM may depend also on the methodology used for aggregate separation from the pooled plasma. While these aggregates appear mostly as globular electrondense complexes when extracted by UC, the SEP methodology seems to generate more fibrous structures. As already mentioned, the use of affinity capture with the SEP binders may cover a broader range of proteins and polymers in different stages of aggregation which end up in the final extraction product. The presence of salt contaminants in the final pellet (Figure 3.7F) may also play a part in the conformation of the macromolecular complexes found in the pellet. Additional TEM analysis will have to be undertaken in order to better define these circulating macromolecular structures. Our preliminary IG-TEM analysis will need to include a range of quality controls, including aggregates generated by heat and sheer stress of Nf proteins and of immunoglobulins/immunocomplexes. Nevertheless, the role of TEM in our study was to confirm the presence of macromolecular formations supporting our enrichment strategy and the data obtained by gel-separation. Further experiments by IG-TEM aimed at obtaining a better characterization of Nf proteins content in NCHs will be needed to optimize this technique for detection and characterization of NCHs. These experiments will have to include quality controls in the form of standard proteins and aggregate formations that can be obtained out of heat or shear stress treatment of protein, antibodies or immunocomplexes (REFERENCE https://microsensbp.com/wp-content/uploads/2015/08/Generation-of-protein-aggregates.pdf and confidential communication by Diagenode). The use of sizing technologies exploiting light diffraction like dynamic light scattering (DLS) or newer tracking methodologies for poly-disperse nanoparticles like nanoparticles tracking analysis (NTA) may also be used in the analysis of NCHs, particularly if conformational changes of these

formations which may have an impact in the disease progression are expected to occur in the development of a pathological state (124).

To conclude, BPP appears to be superior to other extraction methods and it provides higher quality of the data on NCHs acquired by LC-MS/MS. Since MS-based proteomics is the main methodological approach in this project, BPP will be used as the method of choice for NCHs enrichment. This extraction method will support further investigations outlined in this Thesis into NCHs as potential biological substrates of neurodegeneration. A reliable method to extract these macromolecular complexes from biological fluids is essential to further our knowledge on the role of these formations in health and disease. However, the experimental data acquired so far do not answer questions related to the specific origin of the aggregates constituents and about the tissue/fluid environmental conditions which facilitate their formation.

4 Neurofilaments-containing hetero-aggregates (NCHs) in ALS and Healthy Controls plasma samples

4.1 Introduction and aims

The presence of protein aggregates in brain is a hallmark of neurodegeneration (125). However, the assemblement of molecules into macromolecular structures can be observed also in blood and may occur also under non-pathological conditions (126). Circulating aggregates in blood contain proteins that can be relevant as biomarkers of neurodegeneration, such as Neurofilaments (Nf), but as previously shown in Chapter 3, they also include proteins that are involved in physiological responses like inflammation and metabolism, as well as in a number of biochemical changes which are known to occur in defined pathological conditions at a systemic and cellular level.

In this chapter, we aim at providing a biological interpretation of the data outlined in Chapter 3. We will define the changes in Neurofilament-Containing Heteroaggreagates (NCHs) protein composition which occur in a neurodegenerative condition like Amyotrophic Lateral Sclerosis (ALS), using the NCHs proteome of HC as reference. Our goal is also to incorporate in the comparative analysis of protein aggregates those extracted from brain, extending from the preliminary data anticipated in Chapter 3.

To study NCHs and evaluate changes related to the establishment of a pathological state like amyotrophic lateral sclerosis (ALS), pooled plasma sample from ALS patients (ALS-PPS; Table 3.8) and from healthy controls (HC-PPS; Table 3.1) were investigated along with NCHs extracted from brain of patients suffering from ALS. The brain NCHs analysis did not include a comparison with a healthy brain. Nevertheless, brain aggregates composition in ALS was used as the reference to evaluate any molecular feature in circulating NCHs that could mirror brain pathology and be specific of neurodegeneration.

4.2 NfH in pooled plasma samples (PPS)

Western Blot (WB) and immunogold transmission electron microscopy (IG-TEM) data to evaluate the NCHs enrichment process has shown that Neurofilament Heavy (NfH) has a higher tendency to assemble within aggregates than other Nf

proteins including Nf medium chain (NfM) and light chain (NfL). For this reason, NfH was chosen as target Nf isoform for the experiments described below (Figure 4.1).

Unprocessed pooled plasma samples from ALS and HC showed the same NfH bands profile with the highest band at 460 KDa which may represent a NfH dimer in blood (Figure 4.1, lane 9 and 10). When comparing ALS-PPS and HC-PPS products of extraction (final pellets) resolved in urea and PBS it was possible to appreciate a different profile: ALS-PPS and HC-PPS presented NfH bands at 238 KDa and at much higher MW, while only ALS-PPS showed bands at 117 and 60 KDa which may represent NfH fragments (Figure 4.1, lane 1 and 2).

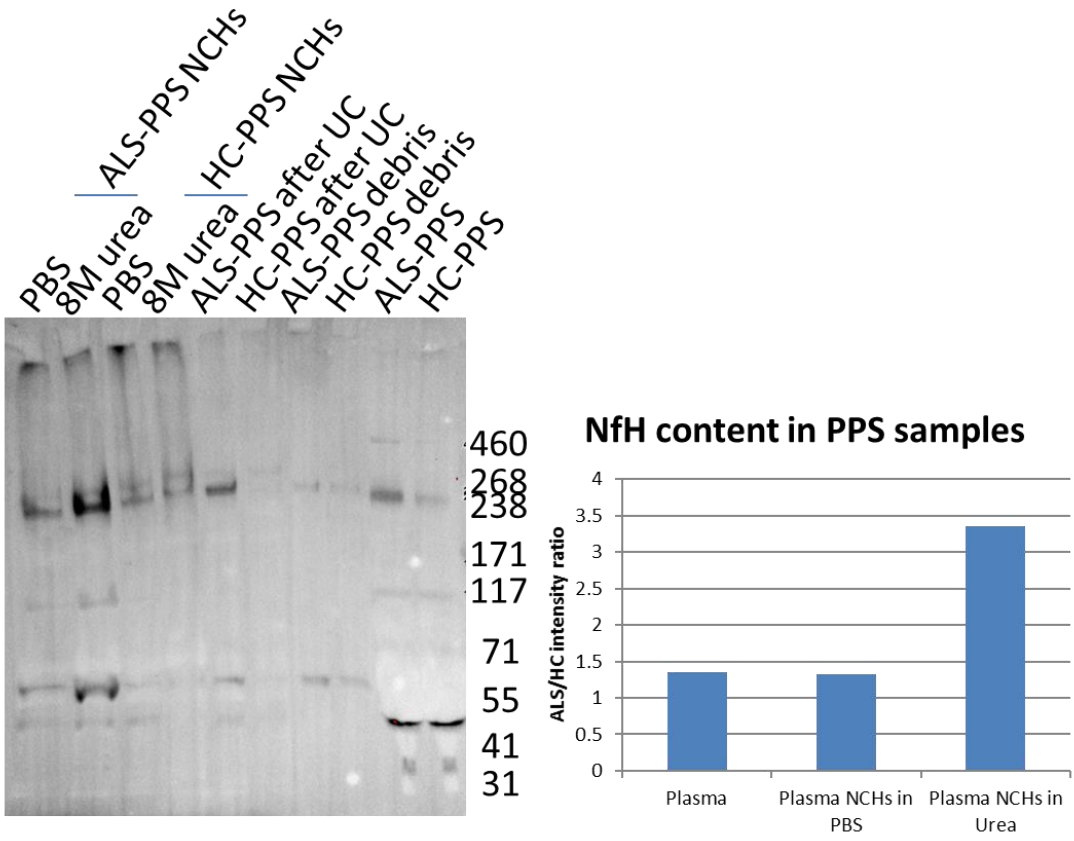


Figure 4.1. Neurofilament Heavy (NfH) content in ALS-Pooled Plasma Sample (ALS-PPS) and Healthy Control-Pooled Plasma Sample (HC-PPS).

NfH Western Blot (WB) (left hand side; already reported in Figure 3.8), showing Neurofilament-Containing Heteroaggregates (NCHs) enriched from HC-PPS and ALS-PPS (lane 1 to 8, from left to right hand side). Lanes 9 and 10 show the NfH content in pooled plasma samples from ALS and HC before enrichment. Lane 7 and 8 show the NfH content in the pellet after the plasma benchtop centrifugation, lanes 5 and 6 show the NfH content in the pellet after the first ultracentrifugation (UC) step, while lanes 1 to 4 show the NfH content in the final NCHs enriched

fractions using two different buffers to resuspend the pellets: PBS (lanes 2 and 4) and 8 M urea (lanes 1 and 3). The different NfH level in ALS-PPS and HC-PPS is shown in the chart (right hand side). The bars represent the ALS/HC ratio of the sum of all the intensities obtained from anti-NfH bands in plasma samples and NCHs enriched fractions lanes treated with both PBS and 8 M urea. ALS/HC ratio is always higher than 1, which indicates that NfH content in ALS-PPS is always higher than HC-PPS in all the fractions under investigation.

The NfH content in ALS-PPS unprocessed plasma was higher than the HC-PPS sample (ALS/HC intensity ratio = 1.35) (Figure 4.1). This observation was in line with previous NfH plasma analysis performed on big ALS and HC cohorts (46,127,128). This datum also supports the use of ALS-PPS and HC-PPS as test-bed for preliminary data acquisition because they represent ALS/HC differences already described in the literature. The same magnitude of difference was shown by the NfH NCHs resuspended in PBS, supporting the idea that plasma NCHs can carry information related to neurology conditions (Figure 4.1). Using 8 M urea to resuspend plasma NCHs increased significantly NfH detection (as described in Paragraph 3.6) as well as the ALS/HC intensity ratio in ALS-PPS sample (Figure 4.1), supporting the hypothesis that NCHs carries valuable biomarkers for diagnosis of neurodegenerative processes. Unfortunately, the use of 8 M urea is not compatible with the chemical conditions required for immunodetection in high throughput techniques such as ELISA.

4.3 Different NCHs composition in ALS and healthy states

As previously reported in Paragraph 3.7.1, NCHs enriched from ALS-PPS and HC-PPS were analysed by Liquid Chromatography coupled with tandem Mass Spectrometry (LC-MS/MS) after in-gel trypsin digestion. Briefly, the two samples were run by electrophoresis in a gradient gel for protein separation, followed by fractionation in ten gel parts for each sample. Each fraction underwent disulphide bonds reduction with dithiothreitol (DTT) and alkylation with iodoacetamide (IAA) and then protein digestion by trypsin. The peptides generated were extracted by dehydration-hydration steps and the samples obtained freeze-dried. Each fraction was then reconstituted and injected into the instrument and protein identification was then performed with Proteome Discoverer 1.4 (PD).

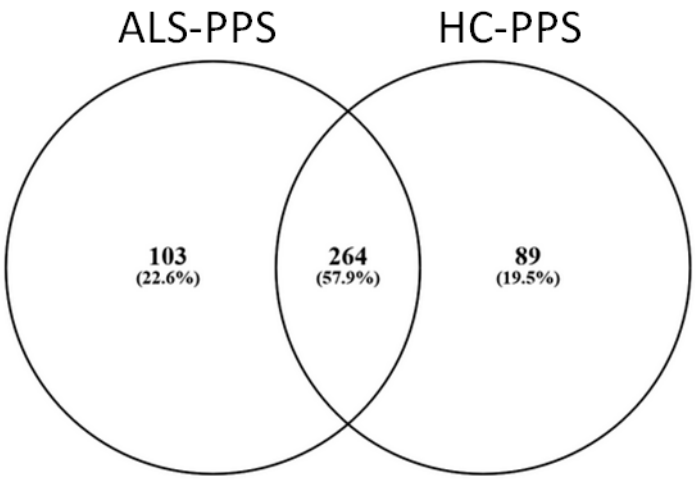


Figure 4.2. Protein groups identified by Proteome Discoverer 1.4 (PD) in NCHs enriched from ALS-PPS and HC-PPS.

The Venn diagram shows the number (with relative percentage of the total number of proteins) of proteins uniquely identified by MS-based proteomics of ALS-PPS and HC-PPS, with the intersection of the two groups displaying the number (and percentage) of proteins identified in both samples. The criteria of peptide selection: minimum peptide confidence: High; minimal number of peptides: 2; count only rank 1 peptides: false, count peptide only in top scored proteins: True.

Following protein identification with PD and filtering out for keratins (contaminant), 367 protein groups were identified in NCHs from ALS-PPS and 363 in NCHs from HC-PPS (Figure 4.2). No Nf proteins were detected in either of the two samples. 264 (57.9% of the total) proteins were shared by the two samples (here defined as shared), while 103 (22.6%) were found only in ALS-PPS (defined as unique ALS) and 89 (19.5%) in HC-PPS (defined as unique Controls) (Figure 4.2). Secondary isoforms (defined by alternative splicing of the same gene) of the three protein lists were generated for unique ALS, unique HC and shared protein groups and then converted into the main isoforms by altering the Uniprot ID (e.g. P02679-2 was converted into P02679) before any functional analysis. This was done because web tools such as Webgestalt or PANTHER do not recognize isoform IDs, effectively introducing a bias in the functional analysis by ignoring part of the information acquired by LC-MS/MS.

We now examine the differences between the ALS and healthy control state in NCHs protein compositions, to pinpoint the biochemical pathways underpinned by the protein aggregation process in ALS-PPS and HC-PPS. In order to define

the biological differences, functional analysis was performed on Webgestalt (108) for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enrichment (Table 4.1, Table 4.2 and Table 4.3).

The ALS-PPS and HC-PPS shared proteome functional analysis showed the enrichment of pathways related to the immune system and its activation (complement and coagulation cascades, staphylococcus aureus infection, pertussis, systemic lupus erythematosus, pathogenic Escherichia coli infection, phagosome and amoebiasis, Table 4.1), while the remaining pathways were mostly linked to cell to cell interaction (ECM-receptor interaction and focal adhesion) and prion diseases. Looking at the statistical significance, FDR values for the top10 pathways were below the usual cut-off of statistical significance (FDR < 0.1) (Table 4.1). Other biochemical pathways were significantly enriched in the shared protein list including vitamin digestion and absorption, malaria, peroxisome proliferator-activated receptors (PPAR) signalling pathway and African trypanosomiasis.

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04610	Complement and coagulation cascades	79	45	1.40	32.06	0.00E+00	0.00E+00
hsa05150	*Staphylococcus aureus* infection	56	15	0.99	15.08	2.07E-14	3.13E-12
hsa05133	Pertussis	76	14	1.35	10.37	4.04E-11	4.08E-09
hsa05020	Prion diseases	35	10	0.62	16.08	2.74E-10	2.08E-08
hsa05322	Systemic lupus erythematosus	135	16	2.40	6.67	1.45E-09	8.80E-08
hsa04512	ECM-receptor interaction	82	11	1.46	7.55	1.72E-07	8.68E-06
hsa05130	Pathogenic *Escherichia coli* infection	55	9	0.98	9.21	4.21E-07	1.82E-05
hsa04145	Phagosome	154	13	2.74	4.75	2.94E-06	1.11E-04
hsa04510	Focal adhesion	203	14	3.61	3.88	1.29E-05	4.33E-04
hsa05146	Amoebiasis	100	9	1.78	5.07	6.44E-05	1.95E-03

Table 4.1. Top10 enriched KEGG pathways identified in the ALS-PPPS and HC-PPS share proteins.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: number of genes included in the given KEGG pathway in Homo sapiens database

Number of genes matched in the category: number of genes identified in the given KEGG pathway

Expected number of genes in the category: calculated as the number of genes submitted divided by the number of the total homo sapiens genes involved in all the KEGG pathway database and multiplied by the number of genes in the category

Ratio of enrichment: calculated as the number of genes matched in the category divided by the expected number of genes in the category

p-Value: statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04512	ECM-receptor interaction	82	8	0.63	12.61	1.72E-07	5.23E-05
hsa04610	Complement and coagulation cascades	79	5	0.61	8.18	3.35E-04	3.49E-02
hsa03050	Proteasome	44	4	0.34	11.75	3.46E-04	3.49E-02
hsa04510	Focal adhesion	203	7	1.57	4.46	8.82E-04	6.68E-02
hsa04145	Phagosome	154	6	1.19	5.04	1.13E-03	6.84E-02
hsa04611	Platelet activation	122	5	0.94	5.30	2.40E-03	1.21E-01
hsa04640	Hematopoietic cell lineage	97	4	0.75	5.33	6.52E-03	2.82E-01
hsa05150	*Staphylococcus aureus* infection	56	3	0.43	6.92	9.09E-03	3.44E-01
hsa00030	Pentose phosphate pathway	30	2	0.23	8.62	2.23E-02	7.49E-01
hsa04810	Regulation of actin cytoskeleton	216	5	1.67	2.99	2.51E-02	7.60E-01

Table 4.2. Top10 enriched KEGG pathways identified in the Unique ALS-PPS protein list.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: number of genes included in the given KEGG pathway in Homo sapiens

Number of genes matched in the category: number of genes identified in the given KEGG pathway

Expected number of genes in the category: calculated as the number of genes submitted divided by the number of the total homo sapiens genes involved in all the KEGG pathway database and multiplied by the number of genes in the category

Ratio of enrichment: calculated as the number of genes matched in the category divided by the expected number of genes in the category

p-Value: statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

KEGG database is generally used for the analysis of groups of genes and of genome information to uncover their functional role in biological systems. It contains information on several organisms and it is extensively used in different areas of research (131). Most of the pathways identified through KEGG analysis of the proteins contained in plasma NCHs from ALS patients are also represented among those identified in the shared protein list (Table 4.1 and Table 4.2), including ECM-receptor interaction, complement and coagulation cascades, focal adhesion, phagosome and *staphylococcus aureus* infection. These are mostly linked to the immune system and its activation (Complement and coagulation cascades, phagosome, platelet activation, hematopoietic cell lineage and staphylococcus aureus infection) but also to cell to cell interaction (ECM-receptor interaction, focal adhesion and regulation of actin cytoskeleton), protein degradation (proteasome) and energetic metabolism (pentose phosphate pathway) (Table 4.2). Importantly, when setting the significance level cut-off at false discovery rate (FDR) < 0.1, the proteasome pathway was the only pathway uniquely enriched in the ALS-PPS and not shared with the HC-PPS (Figure 4.3).

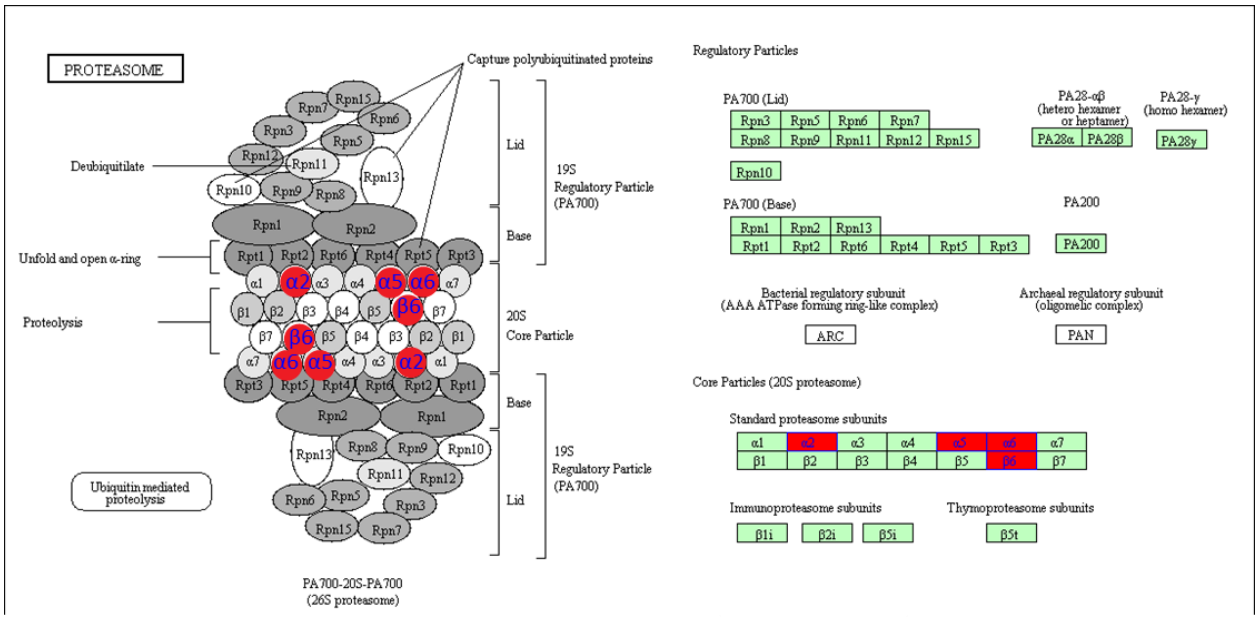


Figure 4.3. Proteasome KEGG pathway (hsa03050).

The figure shows the proteasome KEGG pathway with squared rectangles representing gene products (e.g. proteins). The boxes filled in green indicate the genes/proteins involved in the pathway in Homo sapiens and those in outlined in blue and filled in red are the genes/proteins identified in our ALS-PPS proteome. The subunits of the Core Particles in the 20S proteasome

α2, α5, α6 and β6 are linked to proteasome subunit alpha type-2 (PSMA2), proteasome subunit alpha type-5 (PSMA5), proteasome subunit alpha type-6 (PSMA6) and proteasome subunit beta type-1 (PSMB1), respectively.

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa00010	Glycolysis / Gluconeogenesis	67	7	0.54	13.02	8.73E-07	2.65E-04
hsa01200	Carbon metabolism	114	8	0.91	8.75	2.93E-06	3.43E-04
hsa00030	Pentose phosphate pathway	30	5	0.24	20.77	3.39E-06	3.43E-04
hsa01230	Biosynthesis of amino acids	75	5	0.60	8.31	3.12E-04	2.37E-02
hsa03010	Ribosome	135	5	1.08	4.62	4.36E-03	2.64E-01
hsa05134	Legionellosis	55	3	0.44	6.80	9.56E-03	4.83E-01
hsa04510	Focal adhesion	203	5	1.63	3.07	2.28E-02	8.63E-01
hsa00020	Citrate cycle (TCA cycle)	30	2	0.24	8.31	2.38E-02	8.63E-01
hsa04512	ECM-receptor interaction	82	3	0.66	4.56	2.77E-02	8.63E-01
hsa00051	Fructose and mannose metabolism	33	2	0.26	7.55	2.85E-02	8.63E-01

Table 4.3. Top10 enriched KEGG pathways identified in the Unique HC-PPS protein list.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: number of genes included in the given KEGG pathway in Homo sapiens

Number of genes matched in the category: number of genes identified in the given KEGG pathway

Expected number of genes in the category: calculated as the number of genes submitted divided by the number of the total homo sapiens genes involved in all the KEGG pathway database and multiplied by the number of genes in the category

Ratio of enrichment: calculated as the number of genes matched in the category divided by the expected number of genes in the category

p-Value: statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

Fewer enriched pathways (Table 4.3) were identified in the HC-PPS gene list which were also detected in the shared list (focal adhesion and ECM-receptor interaction (Table 4.1) and in the ALS Unique list (pentose phosphate pathway) (Table 4.2). Most of the enriched pathways found in the HC-PPS gene list were related to energy metabolism, including glycolysis/gluconeogenesis, carbon metabolism, pentose phosphate pathway, TCA cycle and fructose and mannose metabolism, with a smaller number linked to cell to cell interaction like focal adhesion and ECM-receptor interaction, legionellosis, biosynthesis of amino acids and ribosome (Table 4.3). With a significance level set at FDR < 0.1, only glycolysis/gluconeogenesis (Figure 4.4), carbon metabolism (Figure 4.5), pentose phosphate pathway (Figure 4.6) and biosynthesis of amino acids (Figure 4.7) were found to be enriched only in the HC sample.

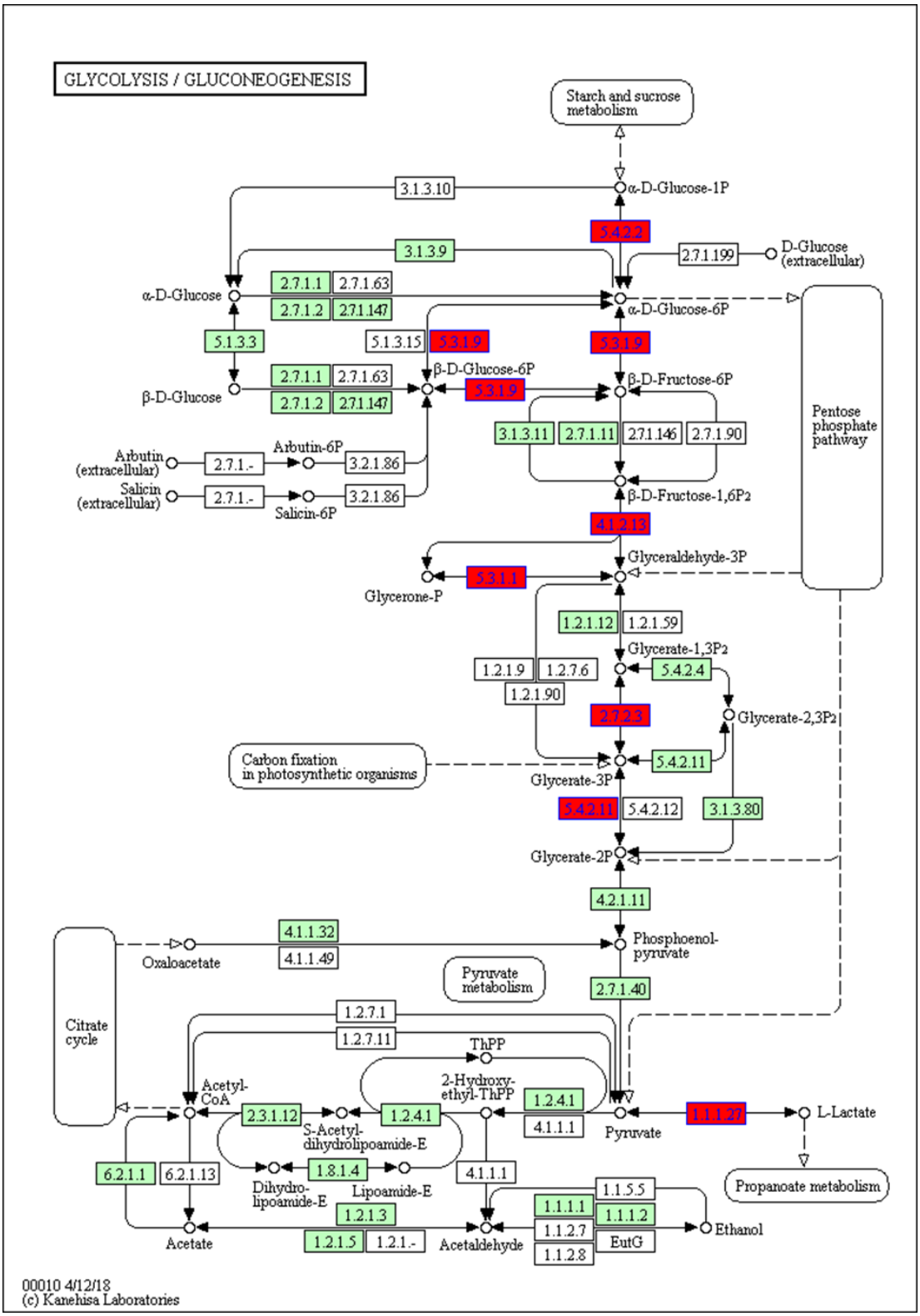


Figure 4.4. Glycolysis/gluconeogenesis KEGG pathway (hsa00010).

The figure shows the Glycolysis/gluconeogenesis pathway with squared rectangles representing gene products (e.g. proteins), rounded rectangles representing maps (e.g. other pathways) and small circles representing chemical compounds. Filled head arrows point to molecular interactions or relations and empty head arrows represent link to or from another map. The objects filled in green indicate the genes/proteins involved in the pathway in Homo sapiens, and those outlined

in blue and filled in red indicate genes/proteins identified in our ALS-PPS proteome. 5.4.2.2, 5.3.1.9, 4.1.2.13, 5.3.1.1, 2.7.2.3, 5.4.2.11, 1.1.1.27 represent phosphoglucomutase-2 (PGM2), glucose-6-phosphate isomerase (GPI), fructose-bisphosphate aldolase A (ALDOA), triosephosphate isomerase (TPI1), phosphoglycerate kinase 1 (PGK1), phosphoglycerate mutase 1 (PGAM1), L-lactate dehydrogenase A chain (LDHA), respectively.

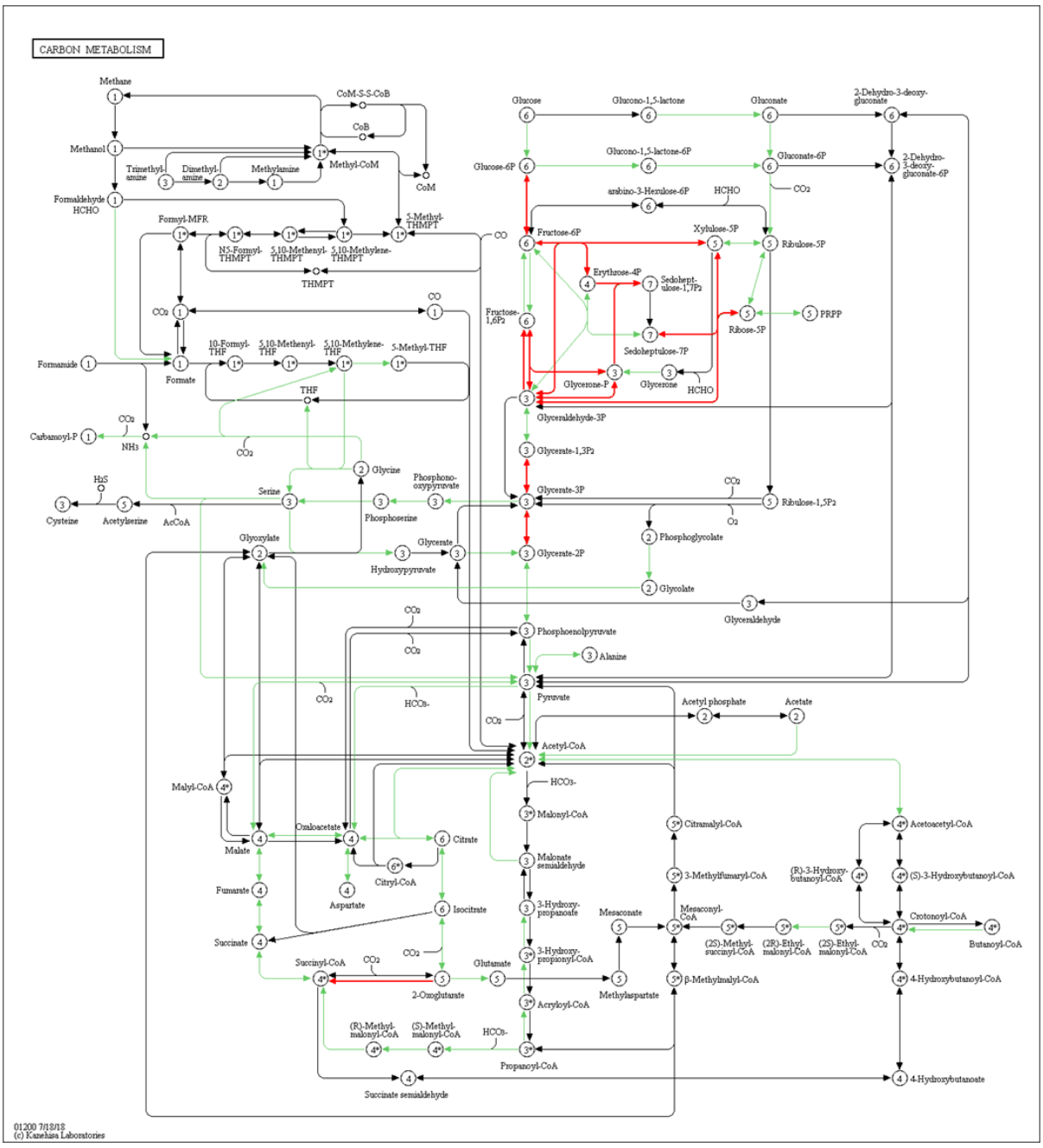


Figure 4.5. Carbon metabolism KEGG pathway (hsa01230).

The figure shows the Carbon metabolism pathway with circles representing the intermediate molecules and the number within the circle representing the number of carbons for each intermediate, while arrows represent the reactions carried out from the enzymes involved in this pathway. The green arrows indicate the enzymes involved in the pathway in Homo sapiens, while the red arrows point to those proteins/enzymes which belong to the HC-PPS unique protein list

(triosephosphate isomerase, TPI1; fructose-bisphosphate aldolase A, ALDOA; phosphoglycerate kinase 1, PGK1; phosphoglycerate mutase 1, PGAM1; dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial, DLST; transketolase, TKT; glucose-6-phosphate isomerase, GPI; 2-oxoglutarate dehydrogenase, mitochondrial, OGDH).

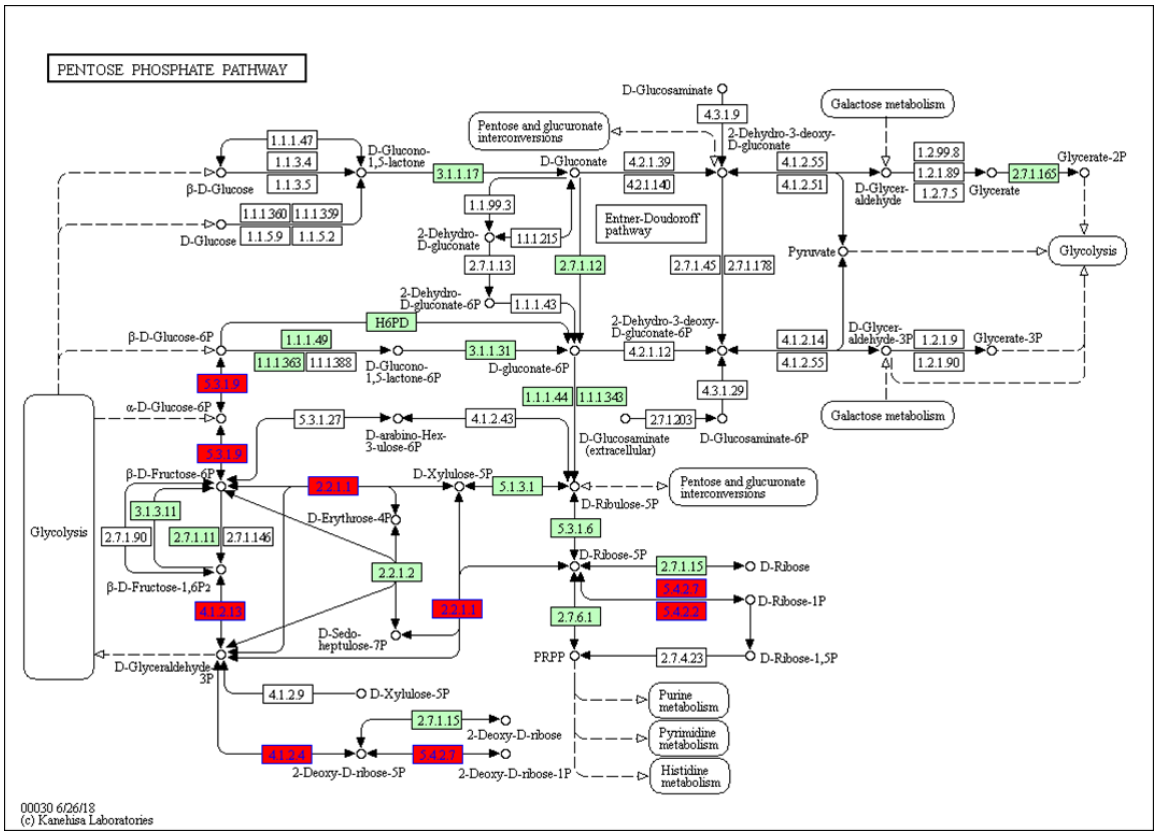


Figure 4.6. Pentose phosphate KEGG pathway (hsa00030).

The figure shows the Pentose phosphate pathway with squared rectangles representing gene products (e.g. proteins), rounded rectangles representing maps (e.g. other pathways), small circles representing chemical compounds, while filled head arrows indicate molecular interaction or relation and empty head arrows the links to or from another map. The objects filled in green indicated the genes/proteins involved in the pathway in Homo sapiens and those outlined in blue and filled in red indicate the genes/proteins found in the HC-PSS proteome. 5.3.1.9, 4.1.2.13, 2.2.1.1, 4.1.2.4, represented glucose-6-phosphate isomerase (GPI), fructose-bisphosphate aldolase A (ALDOA), transketolase (TKT), deoxyribose-phosphate aldolase (DERA), respectively. 5.4.2.2 and 5.4.2.7 were both representing phosphoglucomutase-2 (PGM2).

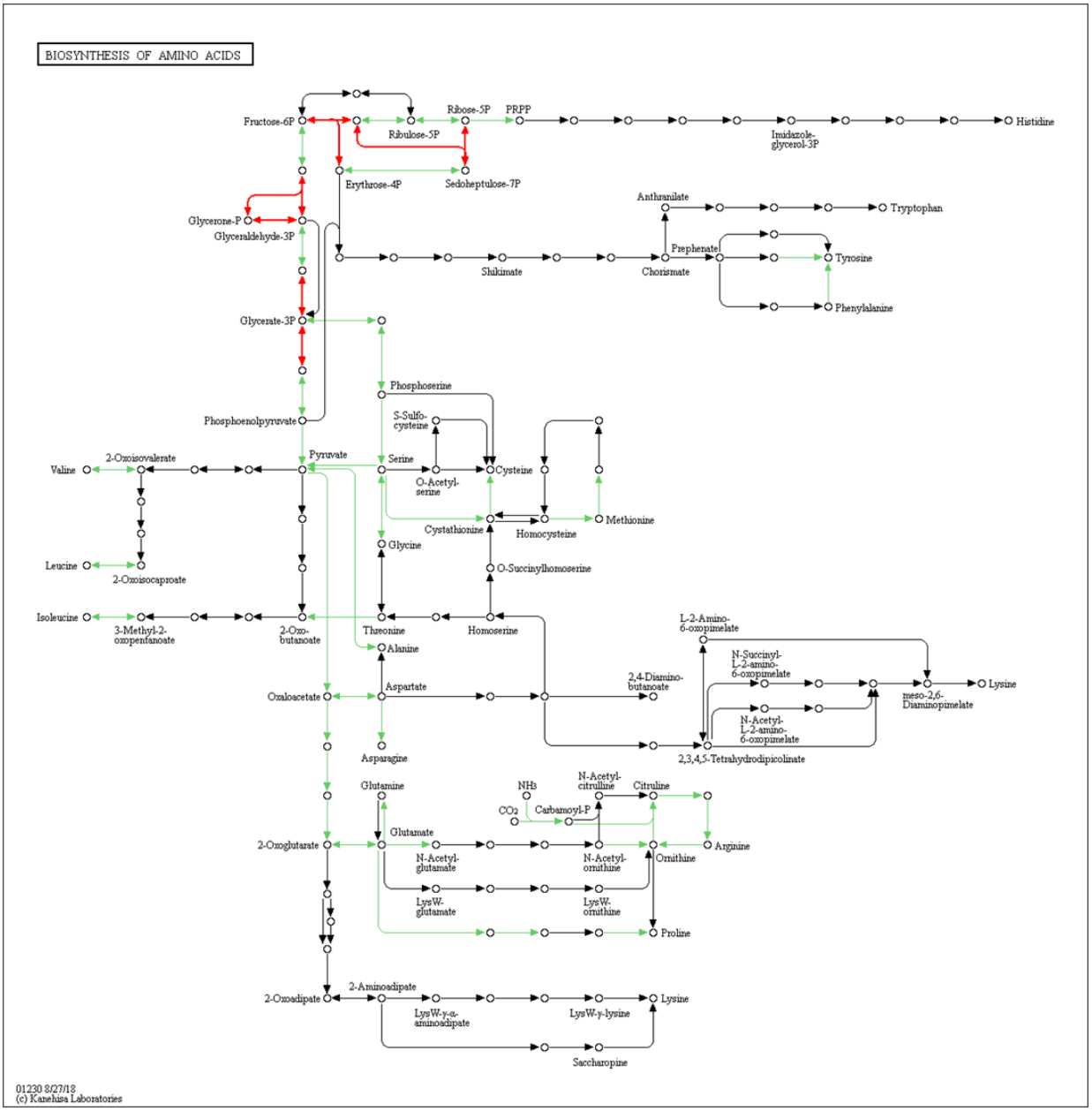


Figure 4.7. Biosynthesis of amino acids KEGG pathway (hsa01200).

The figure shows the Biosynthesis of amino acids pathway with circles representing the intermediate molecules while arrows indicate the reactions carried out by the enzymes involved in this pathway. The green arrows indicate the enzymes involved in the pathway in Homo sapiens, while the red arrows point to the proteins/enzymes identified in the HC-PPS proteome, including triosephosphate isomerase, TPI1; fructose-bisphosphate aldolase A, ALDOA; phosphoglycerate kinase 1, PGK1; phosphoglycerate mutase 1 and PGAM1; transketolase, TKT.

Interestingly, none of the proteins belonging to the “biosynthesis of amino acids” pathway were directly involved in the biosynthesis of amino acids but were mostly key molecules in the production of ribose-5 phosphate and phosphoenolpyruvate, which are also intermediates in the other pathways found to be enriched in our

Unique HC-PPS proteome. This may indicate that the biosynthesis of amino acids pathway is identified in the functional analysis only because of proteins frequently recurring in other pathways and not because of the identification of proteins which are exclusively represented in this pathway.

An enrichment analysis was performed to highlight over-represented functional terms within each of the datasets, taking all human protein-coding genes as the basis and theoretical background for the analysis (132). A specific KEGG pathway was considered enriched if the number of genes represented in that category (*k*) was higher than the expected enrichment value (k_e). This value is calculated by the number of genes in the submitted gene list (n) (e.g. the 103 genes represented in the unique ALS protein list) divided by the number of genes in the reference gene list (m) (e.g. the 6979 genes representing the entire KEGG pathway database) and multiplied by the number of genes represented in a given category within the reference gene list (j) (e.g. the 44 genes involved in the proteasome pathway in the KEGG pathway database).

$$ke = \frac{n}{m} * j$$

Since the gene sets are independent, if k_e is lower than *k* then the category is considered enriched and based on statistical analysis by Fisher's exact test (132).

However, in our analysis the aim was to evaluate if the proteins exclusively identified in one of the two samples were representative of exclusive pathways in one of the two conditions (ALS or HC) rather than a bigger and wider proteome, like the one represented by the KEGG pathway database. For this reason a subsequent enrichment analysis was performed where the entire protein list obtained from the ALS sample was considered as the reference gene set and the unique ALS protein list was the submitted gene list. For this evaluation only the enriched KEGG pathways listed in Table 4.2 with FDR < 0.1 were considered. As a result, the "complement and coagulation cascades" pathway was no longer considered enriched as the number of genes represented in this pathway was lower than the expected value ($k < k_e$). Because in our specific analysis the two gene data sets were not independent, to test the enrichment at a statistical level, we have performed a hypergeometric test (132). The proteasome pathway emerged as the only one with a significant p-value (p=0.028), while ECM-receptor

interaction (p=0.084), focal adhesion (p=0.177) and phagosome (p=0.204) were not statistically significant. The same was applied to the HC data set using the entire protein list as the reference gene set to evaluate the enrichment in the Unique HC list. All pathways with FDR < 0.1 (Table 4.3) were again found enriched, but after significance of enrichment analysis as reported above, only glycolysis/gluconeogenesis (p=0.009), pentose phosphate pathway (p=0.003) and carbon metabolism (p=0.008) were considered significantly enriched, while biosynthesis of amino acids (p=0.08) was not significant. This approach provided a strong basis to consider these four pathways as representative for the difference in NCHs composition between ALS and healthy state.

4.4 LC-MS/MS analysis of Brain NCHs

The methodology for extraction of NCHs from brain has already been detailed in Paragraph 2.2.2.1. Briefly, two brain samples (Brain1 and Brain2) were homogenised and pooled as described in Paragraph 2.1.2.1 and NCHs enrichment performed according to the Brain-Plasma Protocol (BPP; Paragraph 2.2.2.1). The final product of NCHs enrichment and the samples collected during the intermediate steps of the extraction process were evaluated by WB for the presence of Nf proteins (Figure 4.8).

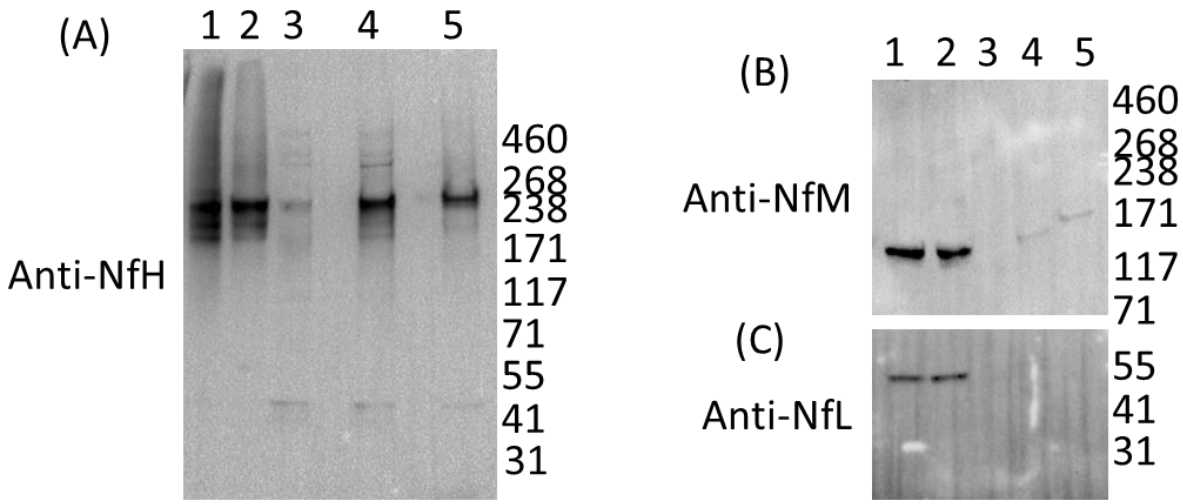


Figure 4.8. Neurofilament (Nf) content in aggregate fractions extracted from Brain examined by Western Blot analysis (WB): neurofilament-containing heteroaggregates (NCHs)

Each fraction obtained using the Brain-Plasma Protocol (BPP) aggregate enrichment protocol and products obtained in the intermediate steps of enrichment were loaded into a gradient gel to

evaluate Nf Heavy (NfH; A), Nf Medium (NfM; B) and Nf Light (NfL; C) protein content. The anti-NfH blot (A) showed the presence of NfH at the expected molecular weight in each fraction (lanes 1 to 5), with the main band at around 238/268 KDa. Anti-NfM blot showed a band between 117 and 171 KDa in lanes 1, 2, 4 and 5, while anti-NfL blot detected NfL bands at approximately 55 Kda only in lanes 1 and 2. Lanes: 1, Brain aggregates enriched fraction; 2, pellet after the first ultracentrifugation (UC) step; 3, supernatant after the first UC step; 4, Brain sample after homogenisation; 5, Brain debris after homogenization as described in 2.1.2.1.

NfH was detected at different MW with the most intense band at the 238/268 KDa (Figure 4.8A). A band was also detected at around 50 KDa in Brain debris (lane 5, Figure 4.8A), in the brain sample (lane 4, Figure 4.8A) and supernatant after the first ultracentrifugation (UC) step of the Brain-Plasma Protocol (BPP, lane 3, Figure 4.8A), but not in the pellet after the first UC and final NCHs enriched fraction (lane 2 and 1; Figure 4.8A). In addition, brain homogenate and brain supernatant after the first UC (Figure 4.8, lane 3 and 4) showed a smear (more evident in lane 4) of the main band and also additional faint bands below 460 KDa, suggesting the presence of several NfH species (e.g. dimers) and complexes or NfH, which may have a different profile of post-translational modifications (PTMs) (Figure 4.8A).

Importantly, the brain aggregates enriched fraction (lane 1) and the pellet after the first UC (lane 2) showed a long smear for high MW formations, confirming the enrichment of neurofilament-containing heteroaggregates (NCHs) in Brain (Figure 4.8A). All Nf protein isoforms seemed to be highly represented in the final fraction enriched with protein aggregates from Brain sample (lane 1 and 2, Figure 4.8). NfM and NfL were detected in the final fractions (lanes 1 and 2, Figure 4.8B and C) in the extraction process but had a very low expression (NfM; lane 3-5, Figure 4.8B) or were totally absent (NfL; lane 3-5, Figure 4.8B) in all intermediate products of the enrichment process. Overall, the analysis of Nf isoforms content in brain indicated, as already shown for circulating NCHs, that unlike NfM and NfL, NfH was present within high molecular weight complexes.

4.4.1 Macromolecular structures in Brain NCHs

In addition to the characterization of NCHs presence in brain by western blot, the same samples were analysed by transmission electron microscopy (TEM) to assess presence of macromolecular structures (Figure 4.9). Because of the high

background obtained in the varoious attempts of IG-TEM in plasma NCHs, immunolabelling was not performed with brain samples but analysis was limited to structure analysis by negative staining.

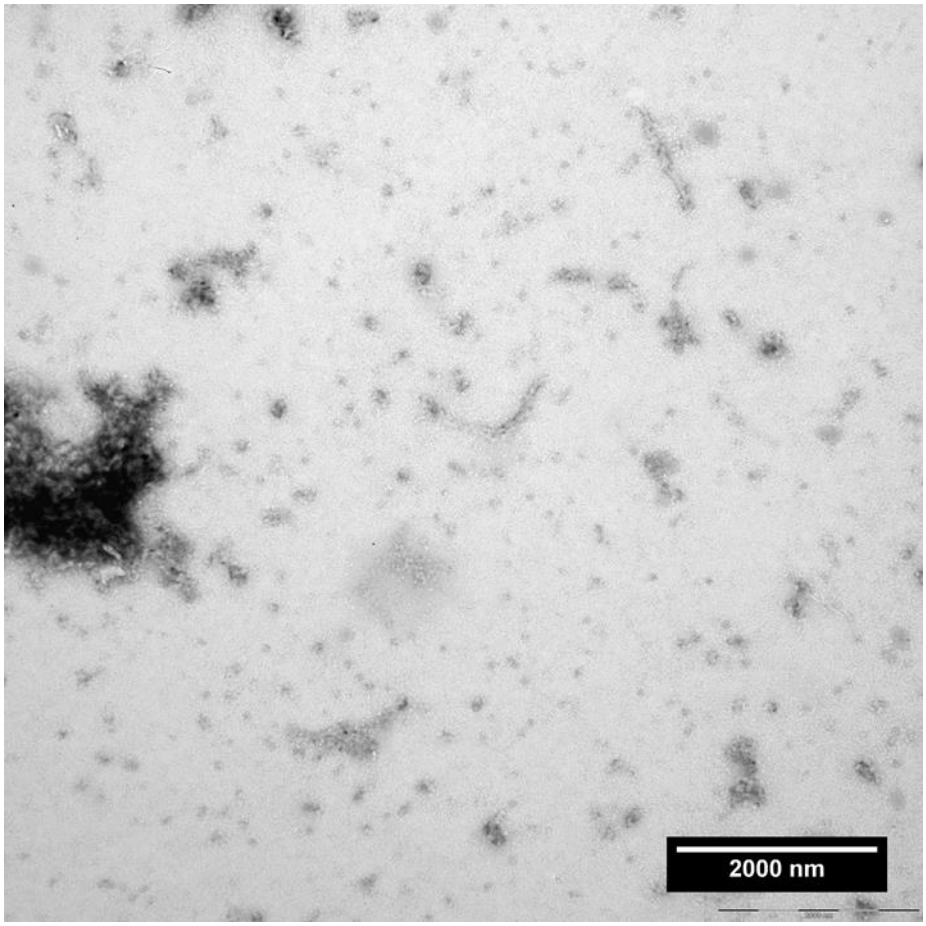


Figure 4.9. TEM analysis of aggregates fraction extracted from brain.
A representative micrograph showing an amorphous electron-dense formation on the left-hand side of the figure. The micrograph also shows the presence of short filamentous and small rounded formations. Scale bar on the low right corner of the figure.

TEM analysis showed the presence of amorphous, globular electron-dense material in the aggregates enriched fraction prepared from ALS brain samples confirming the presence of macromolecular formations. In addition, brain micrograph showed other structures that were not revealed in plasma NCHs (Figure 4.10). These included filamentous shape with a curved appearance and rough-looking surface (Figure 4.10). These filamentous structures had a length ranging from 150 nm to more than 1 µm, while diameter ranged between 10 and 15 nm. The other particles looked donuts-like, rounded with a hole in the middle

(Figure 4.10). These donut-like particles showed a diameter ranging between 20 and 25 nm.

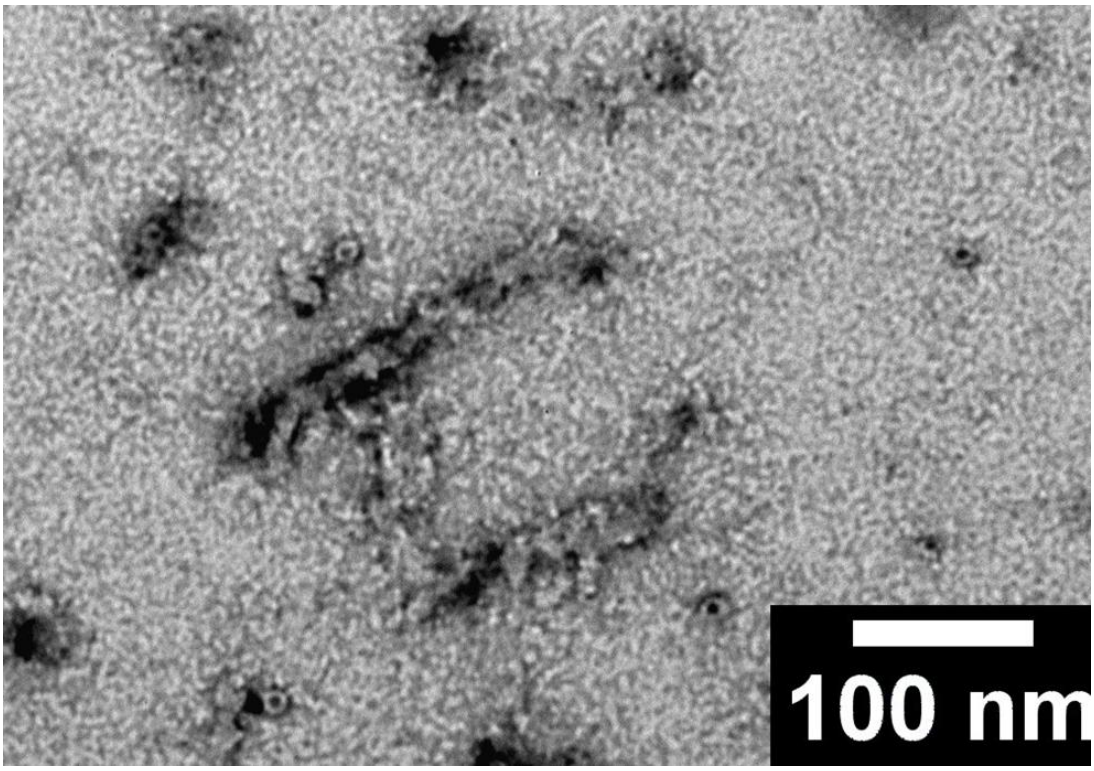


Figure 4.10. Electron-dense formations identified in brain NCHs: filamentous and "donuts"-like appearance.

The micrograph shows details of the filamentous and rounded structures in the aggregates enriched from brain. Scale bar on the low right corner.

4.4.2 Analysis of brain NCHs proteome

Further evaluation of the LC-MS/MS proteome of the aggregate-enriched fraction from ALS brains was undertaken (Paragraph 4.4). Following spectra acquisition, proteins were identified using proteome discoverer 1.4 (PD) and keratin proteins filtered out, generating a final list of 48 protein groups (Table 4.4). Selection criteria were the following: minimum peptide confidence: high; minimum number of peptides: 2, count only rank 1 peptides: false, count peptide only in top scored proteins: true. The three Nf protein isoforms were all identified in the proteome of the brain aggregates fraction. In particular, NfH was identified as isoform 2 with a sequence coverage of 19.52%. NfM had a sequence coverage of 20.85% and NfL 24.49%. Other ALS-relevant proteins such as Superoxide dismutase 1 (SOD1) or TAR DNA-binding protein 43 (TDP-43), known also to be within pathological protein aggregates, were not identified.

Uniprot ID	Protein name	Score	Coverage %	Proteins	Unique Peptides	Peptides	PSMs
O14594	Neurocan core protein	29.64	6.06	1	7	7	13
O15020-2	Isoform 2 of Spectrin beta chain, non-erythrocytic 2	16.61	3.51	2	5	7	7
O43301	Heat shock 70 kDa protein 12A	11.16	8.15	1	4	4	5
O60641-4	Isoform 4 of Clathrin coat assembly protein AP180	46.88	10.38	2	8	8	17
P04350	Tubulin beta-4A chain	65.81	26.35	1	4	10	27
P07196	Neurofilament light polypeptide	65.78	24.49	1	10	13	26
P07197	Neurofilament medium polypeptide	192.52	20.85	1	13	16	85
P08670	Vimentin	8.20	6.44	1	2	3	4
P11137-3	Isoform 3 of Microtubule-associated protein 2	62.12	7.9	2	11	11	22
P12036-2	Isoform 2 of Neurofilament heavy polypeptide	158.61	19.52	2	14	17	66
P12109	Collagen alpha-1(VI) chain	10.13	3.31	1	3	3	4
P12110	Collagen alpha-2(VI) chain	16.05	5.59	1	5	5	7
P12111-4	Isoform 4 of Collagen alpha-3(VI) chain	41.47	4.94	3	12	12	16
P13611-4	Isoform V3 of Versican core protein	42.82	14.35	5	8	8	20
P35579-2	Isoform 2 of Myosin-9	22.36	5.64	2	2	7	10
P35580-3	Isoform 3 of Myosin-10	39.83	6.46	5	5	11	21
P35749-4	Isoform 4 of Myosin-11	26.65	4.85	4	2	9	12

Uniprot ID	Protein name	Score	Coverage %	Proteins	Unique Peptides	Peptides	PSMs
P46379-4	Isoform 4 of Large proline-rich protein BAG6	5.06	1.99	5	2	2	2
P46459-2	Isoform 2 of Vesicle-fusing ATPase	7.07	6.68	2	3	3	4
P60709	Actin, cytoplasmic 1	4.93	7.47	2	2	2	3
P68363	Tubulin alpha-1B chain	71.92	32.82	1	3	11	29
P68366-2	Isoform 2 of Tubulin alpha-4A chain	53.98	31.87	2	2	10	23
Q00610-2	Isoform 2 of Clathrin heavy chain 1	602.89	57.72	2	66	66	282
Q01082	Spectrin beta chain, non-erythrocytic 1	252.70	25.55	1	47	49	115
Q01484	Ankyrin-2	48.98	4.9	1	17	17	23
Q01813-2	Isoform 2 of ATP-dependent 6-phosphofructokinase, platelet type	6.05	2.71	2	2	2	2
Q05193-3	Isoform 3 of Dynamin-1	30.73	10.58	4	8	8	11
Q13813-3	Isoform 3 of Spectrin alpha chain, non-erythrocytic 1	301.11	29.2	3	58	58	130
Q13885	Tubulin beta-2A chain	54.56	24.49	1	4	10	23
Q14204	Cytoplasmic dynein 1 heavy chain 1	117.01	9.54	1	37	37	66
Q15149-7	Isoform 7 of Plectin	115.33	8.99	8	35	35	45
Q16352	Alpha-internexin	46.03	20.64	1	7	10	17

Uniprot ID	Protein name	Score	Coverage %	Proteins	Unique Peptides	Peptides	PSMs
Q6DN90-2	Isoform 2 of IQ motif and SEC7 domain-containing protein 1	4.95	2.14	2	2	2	2
Q7Z4S6-6	Isoform 6 of Kinesin-like protein KIF21A	35.99	5.98	5	8	8	16
Q8TDJ6	DmX-like protein 2	14.34	2.24	2	4	4	5
Q8WXD9	Caskin-1	5.19	2.45	1	2	2	2
Q92752	Tenascin-R	82.10	15.83	1	14	14	36
Q96GW7	Brevican core protein	34.22	9.55	1	8	8	17
Q99250	Sodium channel protein type 2 subunit alpha	10.27	1.75	2	3	3	4
Q99490-2	Isoform 2 of Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 2	4.69	2.63	2	2	2	2
Q9H2M9	Rab3 GTPase-activating protein non-catalytic subunit	5.64	1.22	1	2	2	2
Q9H2X9-2	Isoform 2 of Solute carrier family 12 member 5	31.31	9.77	2	10	10	19
Q9H4G0-4	Isoform 4 of Band 4.1-like protein 1	2.97	4.28	3	2	2	2
Q9UQM7	Calcium/calmodulin-dependent protein kinase type II subunit alpha	8.65	5.23	2	2	2	3
Q9Y2J2-2	Isoform 2 of Band 4.1-like protein 3	14.45	4.05	3	3	3	5

Uniprot ID	Protein name	Score	Coverage %	Proteins	Unique Peptides	Peptides	PSMs
Q9Y4E6-2	Isoform 2 of WD repeat-containing protein 7	8.31	2.06	2	3	3	3
Q9Y4I1-2	Isoform 2 of Unconventional myosin-Va	89.72	13.18	2	19	19	44
Q9Y6D5	Brefeldin A-inhibited guanine nucleotide-exchange protein 2	6.20	1.12	2	2	2	2

Table 4.4. List of proteins identified by LC-MS/MS in brain aggregate fraction using Proteome Discoverer 1.4 (PD).

The table shows proteins identified in brain aggregates by LC-MS/MS and relative detection features (e.g. score). The selection criteria applied were the following: minimum peptide confidence: high; minimal number of peptides: 2, count only rank 1 peptides: false, count peptide only in top scored proteins: true.

Uniprot ID: Uniprot database protein identifier

Protein name: protein full name recommended by Uniprot

Score: sum of the scores computed using the PD SEQUEST-HT algorithm for the individual peptides

Coverage%: percentage of coverage calculated by dividing the number of amino acids in all peptides detected by the total number of amino acids in the entire protein sequence

Proteins: number of proteins identified matching the identified peptides

Unique Peptides: number of peptide sequences unique to a protein group

Peptides: number of distinct peptide sequences in the protein group

PSMs (Peptide Spectrum Matches): total number of identified peptide sequences for the protein including redundant pepetides

4.4.2.1 Functional analysis of the brain aggregates proteome

Protein groups identified in Brain NCHs as secondary isoforms were converted into the main protein isoforms before using Webgestalt (132) for KEGG pathway enrichment analysis (Table 4.5).

The pathways identified as enriched in the Brain NCHs proteome were mostly related to cell to cell interaction (tight junction, ECM-receptor interaction, gap junction, focal adhesion and synaptic vesicle cycle) and to the immune system activation (pathogenic *Escherichia coli* infection, phagosome and salmonella infection), but also to ALS and apoptosis, the only pathway with FDR > 0.1 (Table 4.5). The three Nf isoforms were found only in the ALS pathway and not in the other enriched pathways. Examples of protein represented only in defined pathways included Vesicle-fusing ATPase (NSF), Clathrin heavy chain 1 (CLTC), and Dynamin-1 (DNM1) for the synaptic vesicle cycle pathway and Myosin-11 (MYH11), Band 4.1-like protein 1 (EPB41L1) and Band 4.1-like protein 3 found in the tight junction pathway.

4.4.3 Comparison between plasma and brain NCHs proteomes

The pathway enrichment analysis in brain NCHs showed shared features with the plasma NCHs functional analysis. Four KEGG pathways (pathogenic Escherichia coli infection, phagosome, ECM-receptor interaction and focal adhesion) were represented in both Brain NCHs and the pathways shared by ALS and HC plasma NCHs (Table 4.1 and Table 4.5). However, only few proteins were in common between brain and plasma NCHs (Figure 4.11). Only five proteins (actin cytoplasmic 1, tubulin alpha-4A chain isoform 2, clathrin heavy chain 1 isoform 2, collagen alpha-1(VI) and plectin isoform 7) were shared by Brain, HC and ALS NCHs, while only one protein (cytoplasmic dynein 1 heavy chain 1) was expressed in ALS and Brain NCHs only and one (collagen alpha-2(VI)) between HC and Brain NCHs only.

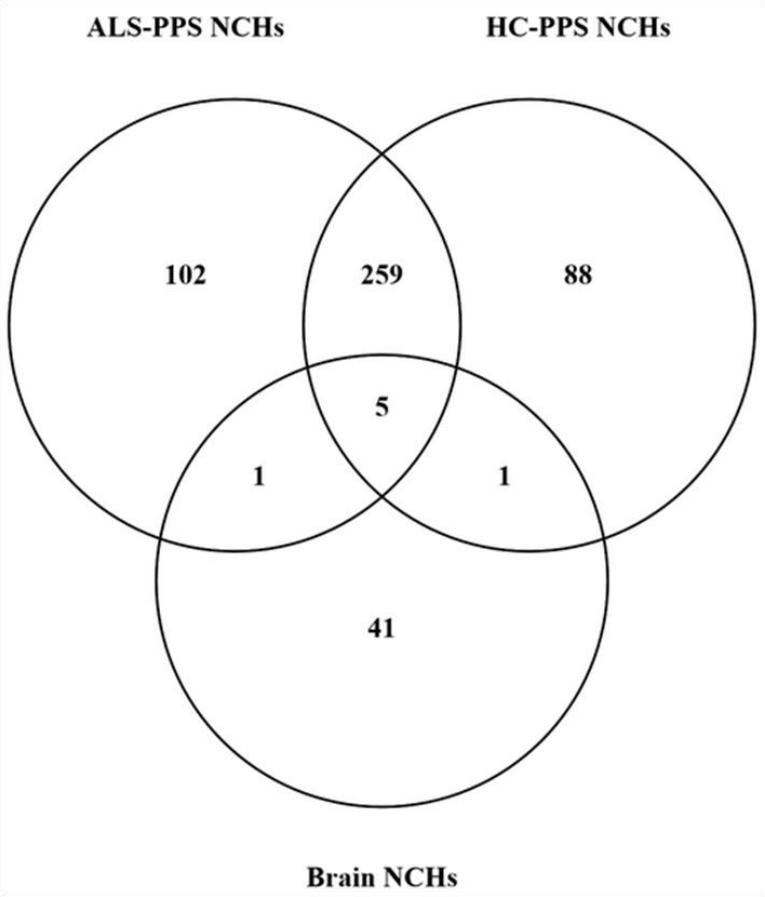


Figure 4.11. ALS-PPS, HC-PPS and Brain NCHs shared protein hits (LC-MS/MS analysis). Venn diagram showing proteins identified by LC-MS/MS analysis and shared by the brain, HC and ALS NCHs. Keratin proteins were filtered out from the lists, while protein isoforms were kept. Five proteins were expressed in all NCHs groups, while Brain NCHs shared only one protein with each of ALS NCHs and HC NCHs.

Gene set	Pathway description	Number of genes in the category	Number of genes matched in the category	Expected number of genes in the category	Ratio of enrichment	p-Value	FDR
hsa04530	Tight junction	139	7	0.64	10.98	2.37E-06	6.55E-04
hsa05130	Pathogenic Escherichia coli infection	55	5	0.25	19.83	4.33E-06	6.55E-04
hsa04145	Phagosome	154	6	0.71	8.50	5.90E-05	5.96E-03
hsa04512	ECM-receptor interaction	82	4	0.38	10.64	4.96E-04	3.27E-02
hsa05132	Salmonella infection	86	4	0.39	10.14	5.94E-04	3.27E-02
hsa04540	Gap junction	88	4	0.40	9.91	6.48E-04	3.27E-02
hsa05014	Amyotrophic lateral sclerosis (ALS)	51	3	0.23	12.83	1.57E-03	6.80E-02
hsa04510	Focal adhesion	203	5	0.93	5.37	2.10E-03	7.97E-02
hsa04721	Synaptic vesicle cycle	63	3	0.29	10.39	2.89E-03	9.71E-02
hsa04210	Apoptosis	140	4	0.64	6.23	3.61E-03	1.09E-01

Table 4.5. Top10 enriched KEGG pathways in the Brain NCHs proteome.

Gene set: KEGG pathway code

Pathway description: KEGG pathway's name

Number of genes in the category: number of genes included in the given KEGG pathway in Homo sapiens

Number of genes matched in the category: number of genes identified in the given KEGG pathway

Expected number of genes in the category: calculated as the number of genes submitted divided by the number of the total Homo sapiens genes included in the KEGG pathway categories and multiplied by the number of genes in the category

Ratio of enrichment: calculated as the number of genes matched in the category divided by the expected number of genes in the category

p-Value: statistical significance calculated by the Fisher's exact test

FDR (False Discovery Rate): calculated from the p-Value adjusted with Benjamini and Hochberg method (109)

4.5 Comparing NCHs data with models of protein aggregation

We have looked at our NCHs proteomic data to investigate molecular mechanism which are reported to influence the susceptibility of proteins to aggregate. To this end, we have used datasets from existing *in vivo* and *in silico* models which are based on data acquired studying yeast models and prion proteins (133–138).

4.5.1 Algorithms for modelling aggregation tendency

An and Harrison (135) identified possible human and yeast-prion-like proteins by different computational methods, including the lowest-probability subsequence (LPS) (139), the Prion Aggregation Prediction Algorithm (PAPA) (140) and prion-like amino acid composition (PLAAC) (141). Using these tools, they have generated a list of prion-like proteins in common between human and yeasts. We have compared this prion-like protein list with the lists of proteins identified in the plasma and brain NCHs proteomes (Figure 4.12).

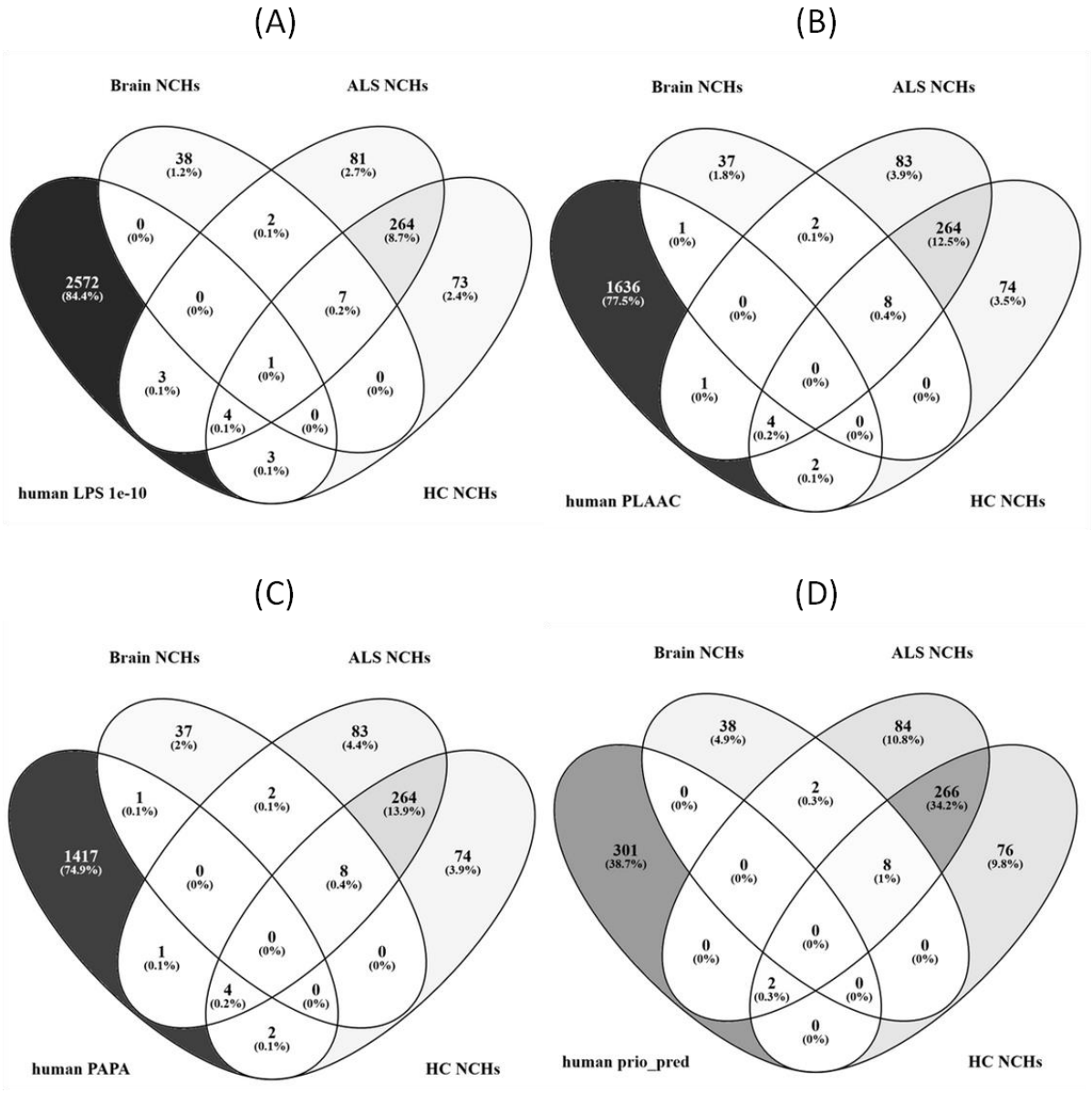


Figure 4.12. Overlap between proteins identified in Brain and plasma NCHs and predicted human prion proteins.

ALS Brain NCHs, ALS and HC plasma NCHs protein lists obtained by LC-MS/MS were compared with (A) the human lowest-probability subsequence (LPS) 1e-10 representing the list of human prion-like proteins computed by LPS (p-value threshold of 1e-10), (B) the human prion-like amino acid composition (PLAAC) list of human prion-like proteins computed by PLAAC, (C) the human Prion Aggregation Prediction Algorithm (PAPA) list of human prion-like proteins computed by PAPA and (D) the human yeast-prion-like proteins as reported by An and Harrison (135). The different grey scale is proportional to the percentage of proteins included in each diagramm intersection. Keratin proteins were filtered out and secondary isoforms were converted into main isoforms.

The final list of human yeast-prion-like proteins reported by An and Harrison is significantly enriched in cytoskeletal proteins (135). A good proportion of the

proteins we have identified in the Brain NCHs proteome were also of cytoskeletal proteins (20% of the total) were included in the cytoskeleton cellular component category identified by PANTHER database. However, the overlap between computed human prion-like proteins and those we have identified in the ALS brain NCHs, and in the ALS and HC plasma NCHs was very poor (Figure 4.12). Eleven proteins were shared between the human LPS list (Figure 4.12A) and the plasma/brain NCHs proteins, where one protein was shared among all the lists (Plectin), three proteins were shared between LPS and ALS NCHs (Cornulin, Semenogelin-1 and Ribosome-binding protein 1), three between LPS and HC NCHs (Envoplakin, Involucrin and Periplakin), and four between ALS and HC NCHs (Desmoplakin, Apolipoprotein A-IV, Hornerin and Filaggrin-2). This level of comparability was also seen when the human PLAAC and the PAPA lists of human prion-like proteins were used (Figure 4.12B), which shared one protein only with Brain NCHs (Neurofilament Light), one with plasma ALS NCHs (Myc target protein 1), two with plasma HC NCHs (ATP-dependent RNS helicase and Alpha-amylase 1) and four with plasma ALS and HC NCHs together (Hornerin, Filaggrin-2, Complement component C7 and Coagulation factor V; Figure 4.12C). Using the final list of human and yeast prion-like proteins produced by An and Harrison (135) (Figure 4.12D), only two of these prion-like proteins were found in (Hornerin and Filaggrin-2) in our plasma ALS and HC NCHs lists.

4.5.2 Yeast model for protein aggregation

In the attempt to acquire a better understanding of mechanisms involved in protein aggregation, Weids, Ibstedt et al. studied protein aggregation in yeast under different type of stress and the associated physicochemical properties, such as protein size, isoelectric point (pI) and hydrophobicity, all considered to be involved in the aggregation process (136). Using the same approach, we have studied the proteins identified in the ALS Brain, ALS and HC plasma NCHs proteomes to define their protein size (obtained from Uniprot) (Figure 4.13), pI (obtained by ExPASy, Compute pI/Mw web tool (142)) (Figure 4.14) and hydrophobicity by the GRAVY (average hydrophobicity and hydrophilicity) score (143) using the GRAVY Calculator (http://www.gravy-calculator.de) (Figure 4.15). This analysis was undertaken using the entire Uniprot human proteome (reviewed sequences only) as reference, which is also the database used for protein identification with Proteome Discoverer v1.4 (PD).

Looking at MW distribution (Figure 4.13), the HC NCHs plasma proteins and those shared between ALS and HC NCHs showed a similar distribution compared to the Human proteome, while the ALS NCHs plasma proteins were clearly more distributed towards higher MW. The ALS Brain NCHs proteins group had a shape similar to the Human proteome but, as seen for the ALS NCHs plasma proteins, they appeared shifted to the right (higher MW). Overall, these data showed a tendence for higher MW size for those proteins identified in the ALS NCHs (both brain and plasma) compared to the entire human proteome, while the HC NCHs plasma proteins showed a lower average protein size (ALS 1.76, HC 1.61, Shared 1.71, Brain 2.10, Human proteome 1.66 - KDa log10). The only statistically significant shift using the human proteome as reference was for the ALS Brain NCHs proteins (unpaired Mann-Whitney-Wilcoxon U-tests and Holm-Bonferroni p-value adjustment, 9.38e-16).

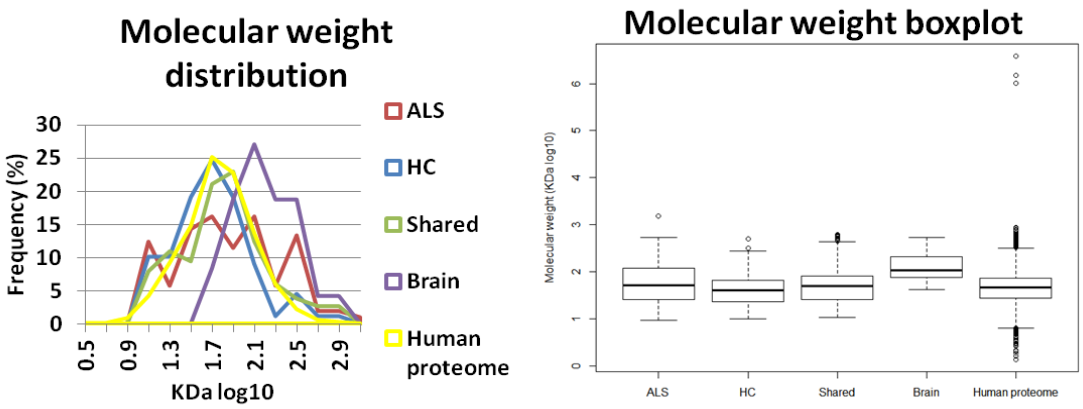


Figure 4.13. Molecular weight distribution of the proteins identified in the plasma and brain NCHs by LC-MS/MS and of the human proteome.

The graphic on the left-hand side shows the distribution of the average protein molecular weights (KDa log10) of the HC, ALS NCHs plasma proteins, of the ALS brain NCHs proteins with reference to the whole genome protein MW distribution. The boxplot (right-hand side) shows the median, first and third quartiles, upper and lower limits, and the points beyond (empty circles) the limits. ALS: NCHs Unique ALS protein list; HC: NCHs Unique HC protein list; Shared: NCHs Shared protein list; Brain: Brain NCHs protein list; Human proteome: Homo Sapiens Uniprot reviewed protein list.

With regards to the isoelectric point, proteins identified in the HC/ALS plasma NCHs distributed differently from the human proteome. The ALS brain NCHs proteins showed a narrower pI distribution compared to the proteins identified in plasma NCHs (Figure 4.14). Overall, the proteins identified in the plasma NCHs

showed a consistently and statistically significant lower pI compared to the human proteome (unpaired Mann-Whitney-Wilcoxon U-tests and Holm-Bonferroni p-value adjustment, 5.93e-12, 1.15e-9, <8.80e-16, 2.70e-10, respectively for ALS, HC, Shared and Brain protein lists) (Figure 4.14).

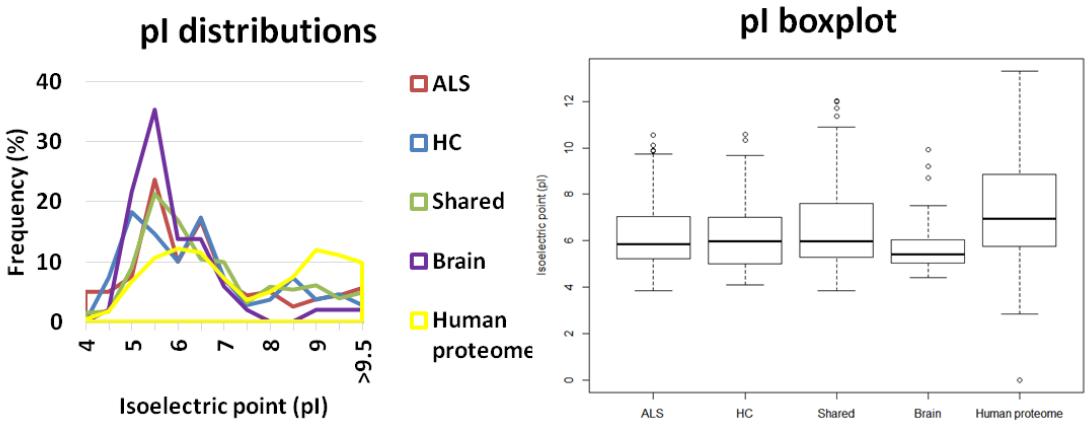


Figure 4.14. Isoelectric point (pI) distribution of the proteins identified in the plasma and brain NCHs by LC-MS/MS and of the human proteome.

The graphic on the left-hand side shows the distribution of pIs of the HC, ALS NCHs plasma proteins, and of the ALS brain NCHs proteins with reference to the whole genome. The boxplot (right-hand side) showed the median, first and third quartiles, upper and lower limits, and the points beyond (empty circles) the limits. ALS: NCHs Unique ALS protein list; HC: NCHs Unique HC protein list; Shared: NCHs Shared protein list; Brain: Brain NCHs protein list; Human proteome: Homo Sapiens Uniprot reviewed protein list.

Observing the distribution of the hydrophobicity measured as GRAVY index, the protein groups generated by the proteomic analysis showed a similar distribution, with similar average values (ALS -0.34, HC -0.33, Shared -0.35, Human proteome -0.34) (Figure 4.15). The only exception was represented by the ALS Brain NCHs protein list which showed statistically significant lower GRAVY values (average -0.50; unpaired Mann-Whitney-Wilcoxon U-tests and Holm-Bonferroni p-value adjustment, 0.03) (Figure 4.15).

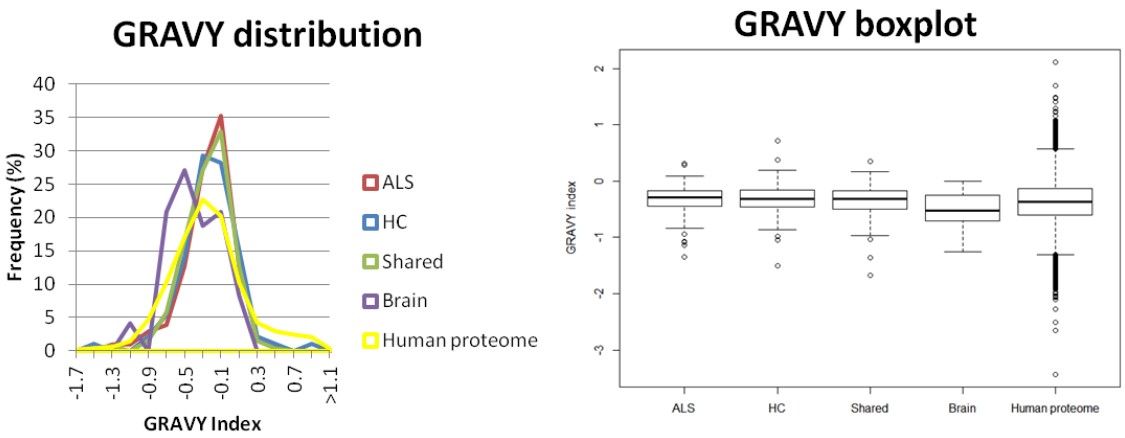


Figure 4.15. GRAVY index distribution of the proteins identified in the plasma and brain NCHs by LC-MS/MS and of the human proteome.

The graphic on the left-hand side shows the distribution of hydrophobicity, as GRAVY index, of the HC, ALS NCHs plasma proteins, and of the ALS brain NCHs proteins with reference to the whole genome. The boxplot (right-hand side) shows the median, first and third quartiles, upper and lower limits, and the points beyond (empty circles) the limits. ALS: NCHs Unique ALS protein list; HC: NCHs Unique HC protein list; Shared: NCHs Shared protein list; Brain: Brain NCHs protein list; Human proteome: Homo Sapiens Uniprot reviewed protein list.

4.6 NCHs resistance to protease digestion with trypsin

In neurodegenerative diseases, it is thought that disordered proteins are prone to form aggregates, behaving in a prion-like manner. This concept arises from the observation that neurodegenerative diseases including ALS, may follow a similar pattern of progression of the disease due to propagation of the pathology in the affected tissues. This progression from an anatomic area to a contiguous one seen in different neurological conditions recalls the way prion proteins diffuse, extending the hallmarks of neurodegeneration to contiguous areas (100,144,145). Another important characteristic of the biochemical behavior of prion proteins is the resistance to cleavage by protease K and trypsin (146), which in turn may be at the basis of prions proteins enhanced tendency to form protein aggregates. It is therefore possible to postulate that protein aggregates in general may include proteins manifesting the same or various degrees of protease resistance and that this biochemical characteristic may be prominent when protein aggregation occurs in a pathological state like for example ALS. In order to confirm whether altered protease resistance could also be found in the NCHs

and be correlated to ALS pathology, digestion with trypsin was performed on brain and plasma NCHs (Figure 4.16).

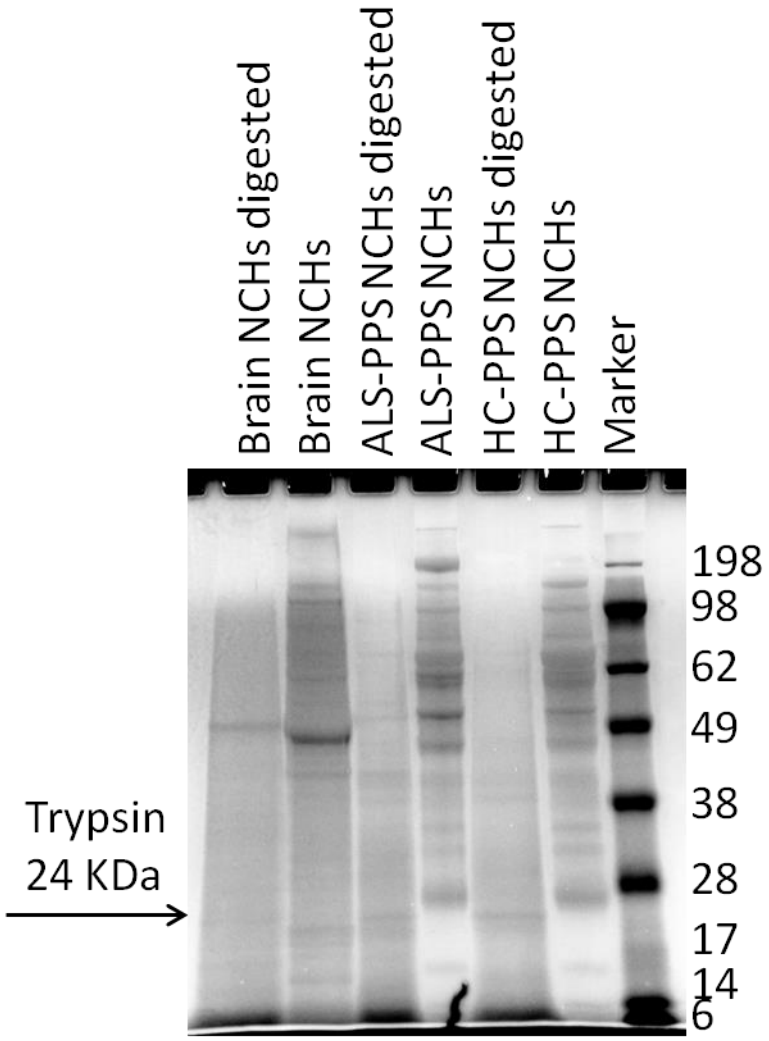


Figure 4.16. Neurofilament-containing heteroaggregates (NCHs) before and after digestion with Trypsin.

NCHs extracted from brain, ALS-PPS and HC-PPS were digested overnight with trypsin. NCHs proteins show a complete and homogeneous trypsin digestion across plasma and brain subgroup. The arrow on the left-hand side highlights the presence of the retained protease enzyme used for aggregates digestion (rather than a digestion fragment). NCHs: neurofilament-containing heteroaggregates (NCHs); ALS-PPS: amyotrophic lateral sclerosis pooled plasma sample; HC-PPS: healthy control pooled plasma sample.

After ALS brain NCHs digestion, it was possible to detect a defined band at 49 KDa, whereas the rest of the plasma NCHs (HC-PP and ALS-PP) digested samples showed a very faint smear going from 98 KDa MW to the bottom of the gel, suggesting a degree of resistance to digestion with trypsin (Figure 4.16). As

discussed later on, the presence of a degree of protease resistance may influence the proteomic results and affect negatively the detection of other ALS relevant proteins like TDP-43 and SOD1.

In order to evaluate protein-specific resistance, WB against NfH (Figure 4.17), TDP-43 (Figure 4.18) and ubiquitinated proteins (Figure 4.19) was also performed on the same brain and plasma-derived aggregates from ALS and HC.

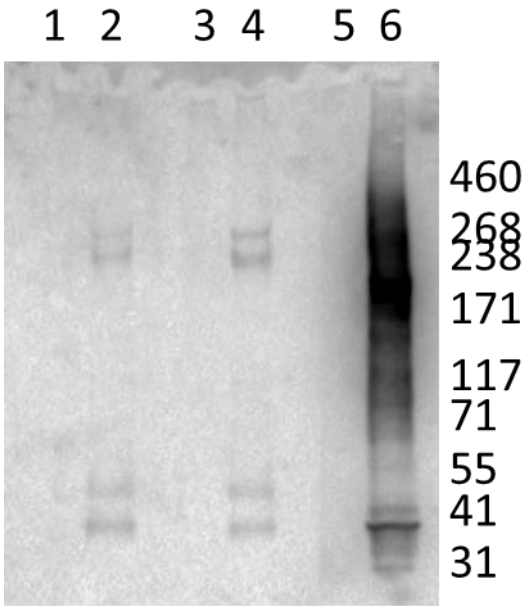


Figure 4.17. Western blot analysis of NfH in brain and plasma NCHs digestion with trypsin. The WB shows the NfH content in ALS-PPS (lane 1 and 2), HC-PPS (lane 3 and 4) and Brain NCHs (lane 5and 6) and the effect of trypsin digestion. The blot indicates that trypsin entirely digests NfH contained in NCHs. Brain NCHs profile is different from plasma, with a significantly different expression of bands above and below the expected molecular weight. Brain NCHs presented a long smear from 268 to 117 KDa, with the main band below 238 KDa and another distinct band between at 41 KDa. Plasma NCHs showed two bands at around 238-268 KDa and other bands at 55 and 41 KDa. Lanes: 1, ALS-PPS NCHs digested with trypsin; 2, undigested ALS-PPS NCHs; 3, HC-PPS NCHs digested with trypsin; 4, undigested HC-PPS; 5, ALS Brain NCHs digested with trypsin; 6, undigested brain NCHs.

NfH did not show any resistance to trypsin protease digestion in either ALS brain and HC/ALS-derived plasma NCHs (Figure 4.17). This refutes our initial hypothesis that NfH could not be detected by MS because of intrinsic resistance to trypsin digestion. ALS Brain and plasma NCHs showed a different WB pattern with regards to NfH content. NfH was detected in plasma samples in four different bands at defined MWs (268, 238 and between 55 and 41 KDa), while Brain NCHs

WB analysis of NfH appeared more as a smear around the levels of more defined MW bands, including from 268 to 117 KDa and at lower MW at around 41 KDa (Figure 4.17). One potential explanation for the reported NfH distribution in Brain could be a higher level of post-translational modifications (PTMs), which may result in a broad number of NfH species detected at different MW (147,148). This could also be true for NfH circulating in blood, where the presence of defined species identified by WB may also relate to the PTM profile.

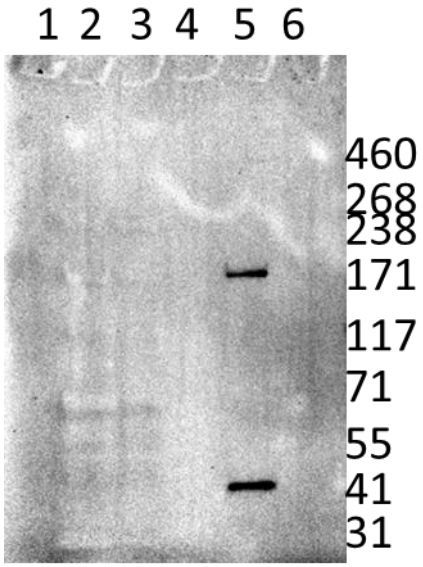


Figure 4.18. Western blot analysis of TDP-43 in brain and plasma NCHs before and after digestion with trypsin.

The WB showed the TDP-43 content in HC-PPS, ALS-PPs and Brain NCHs and the effect of trypsin digestion. The blot shows absence of TDP-43 in NCHs obtained from plasma samples, while TDP-43 in Brain NCHs at 43 and 171 KDa. Brain NCHs digested with trypsin did not show any positive band for TDP-43, against any evidence of protease resistance. Lanes: 1, undigested HC-PPS; 2, HC-PPS NCHs digested with trypsin; 3, undigested ALS-PPS NCHs; 4, ALS-PPS NCHs digested with trypsin; 5, undigested brain NCHs; 6, brain NCHs digested with trypsin.

The two separate brain samples (here referred to as Brain1 and Brain2) which were pooled for the purpose of the experiments detailed in this Thesis, were obtained from the Netherlands Brain Biobank and were both reported to be positive for TDP-43 inclusions after post-mortem analysis. Western blot analysis of TDP-43 in NCHs extracted from brain confirmed the presence of this protein in the Brain at the expected (43 KDa) and higher (171 KDa) MW, suggesting the potential inclusion of this protein in brain aggregates and confirming the quality of the enrichment method we have applied (Figure 4.18). TDP-43 was not

detected in plasma NCHs from HC and ALS patients (Figure 4.18). In addition, NCHs digestion with trypsin did not show any TDP-43 resistance to digestion. This may indicate that failure to detect TDP-43 by MS in our proteomic experiment is not related to resistance to trypsin digestion, the proteolytic enzyme used in the MS procedure, but rather to the presence of PTMs which may interfere with TDP-43 detection with the proteomic pipeline setting utilized in our experiments (149).

Misfolded proteins can be degraded within cells via the proteasome system or through other proteolytic pathways within the lysosomes (150,151). Ubiquitin plays a central role in all these pathways and it is also involved in protein aggregates removal (152–155). ALS and other neurodegenerative diseases where protein aggregation is a major pathological hallmark, such as Alzheimer's, Parkinson's and Huntington's disease, are known to show presence of ubiquitinated proteins within protein aggregates in affected tissues (156–158). To evaluate the degree of protein ubiquitination in the NCHs from HC-PPS, ALS-PPS and in ALS Brain, we have performed a WB against ubiquitinated proteins (Figure 4.19).

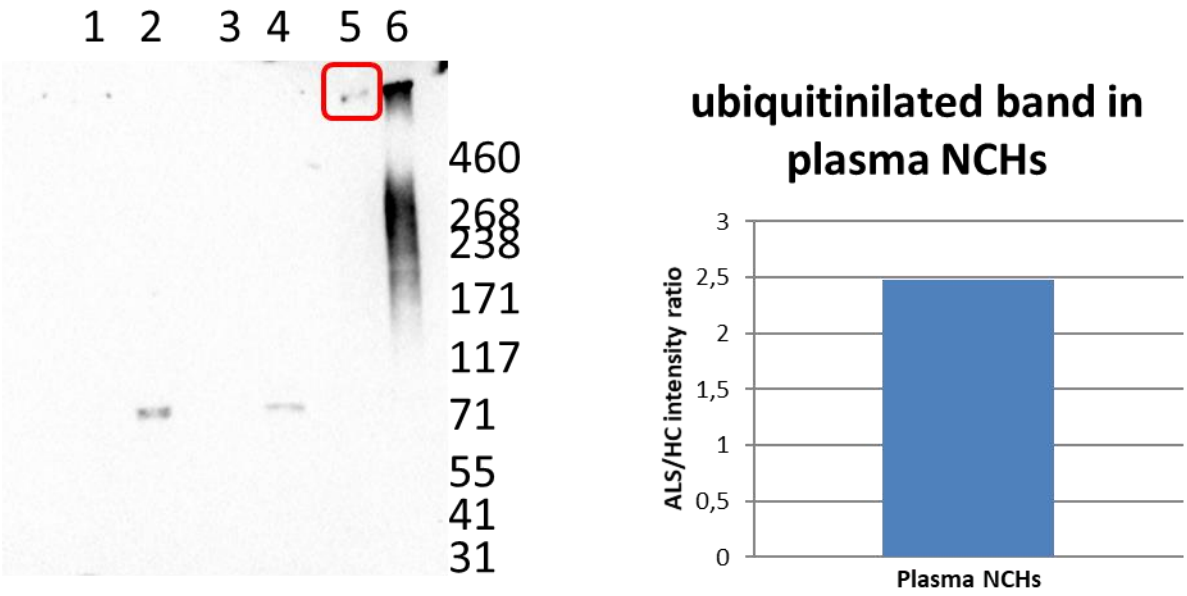


Figure 4.19. Western blot analysis of ubiquitinated proteins in brain and plasma NCHs after digestion with trypsin.

Western Blot (WB, left-hand side) analysis shows the presence of ubiquitinated proteins and the effect of trypsin digestion in NCHs from ALS-PPS (lane 1 and 2), from HC-PPS (lane 3 and 4) and from ALS Brain (lane 5 and 6). Plasma NCHs (lanes 1-4) show ubiquitinated proteins at a

MW of 71 KDa, while Brain NCHs (lane 5 and 6) ubiquitinated proteins were expressed at a MW range between 268 KDa to 171 KDa and at higher MW (above 460 KDa). In the samples we have examined, only ALS Brain NCHs show ubiquitinated proteins resistance to trypsin digestion (high MW in lane 6, highlighted by the red box). The chart (right-hand side) shows the densitometric analysis of the bands containing ubiquitinated proteins found in ALS-PPS (lane 2) and HC-PPS NCHs (lane 4). Lanes: 1, ALS-PPS NCHs digested with trypsin; 2, undigested ALS-PPS NCHs; 3, HC-PPS NCHs digested with trypsin; 4, undigested HC-PPS; 5, Brain NCHs digested with trypsin; 6, Brain NCHs undigested.

Ubiquitinated proteins were detected in NCHs extracted from ALS-PPS, HC-PPS and Brain. Only Brain showed ubiquitinated proteins resistance to trypsin digestion (high molecular weight band, lane 5, Figure 4.19), while NCHs ubiquitinated proteins from ALS-PPS and HC-PPS showed no resistance to trypsin digestion (Figure 4.19). The WB analysis of ubiquitinated proteins in plasma NCHs showed a band at 71 KDa, with the NCHs from ALS-PPS displaying more than the double of the intensity compared to HC-PPS (Figure 4.19). Brain NCHs showed ubiquitinated proteins at a high MW (above 460 KDa) and a smeared band going from above 268 KDa to 171 KDa. WB analysis of ubiquitinated proteins in brain NCHs digested with trypsin showed also a band at high MW suggesting a degree of resistance for a portion of ubiquitinated aggregates of higher molecular weight (Figure 4.19) (159,160). Taken together, the results of this trypsin digestion experiment would suggest that ubiquitin could be co-localised with NfH and TDP43 in the same aggregates (Figure 4.17-Figure 4.19). Therefore, the inclusion of ubiquitination as PTM would probably increase peptide and protein identification after LC-MS/MS analysis through the identification of this specific PTM.

4.7 Conclusion

The work highlighted in this Chapter relates to the characterisation of NCHs, extracted from plasma of HC and ALS patients and from ALS brains. We have looked at NCHs protein composition by MS and at their biochemical properties testing their resistance to trypsin protease digestion. While the information we have gathered represent an initial step towards new and potentially informative biomarkers of neurodegeneration, our preliminary findings require further refinement and validation. So far, we have opted for an investigation of pooled samples coming from patients who differed in terms of age of onset of the disease

and rate of disease progression. The work was so far conducted on pooled plasma samples missing any specific molecular feature linked to each affected individual' phenotypic aspect and proteomic analysis did not provide any quantitative determination and information about specific candidate's regulation in the disease state. In the comparative analysis of plasma NCHs with brain aggregates, we have studied the expression of biological features in association with ALS and how protein profiles of aggregate fractions in brain and plasma compare. In addition, we have so far looked at how the specifics of the proteomes we have acquired depend on the experimental set-up, which include the specific MS workflow and the extraction method.

NfH has already been described as an important biomarker for neurodegeneration in ALS (46,127,128). Other studies highlighted potential problems in the detection of this neurofilament isoform using immunodetection methods like ELISA, due to its propensity to form immunocomplexes or to assemble in other protein aggregates (53). The solubilisation of NCHs derived from plasma, which was carried out with 8 M urea, showed that NfH is present at higher levels in NCHs from ALS patients than from HC, similarly to what seen in plasma NCHs solubilised with PBS (Figure 4.1). The higher NfH expression in ALS NCHs compared to HC NCHs showed in Figure 4.1 suggests that the NfH included in protein aggregates could be more informative as biomarker of neurodegeneration than simple NfH measurements in plasma. On the other hand, solubilisation with 8 M urea may not be compatible with antibody-based high throughput techniques, such as ELISA, reducing the possible application of this finding for large-scale biomarkers analysis. Further investigation including larger cohort of ALS and control individuals may be needed to unravel the real potential of aggregate-phase NfH as biomarker and its applicability to clinical practice.

Electron microscopy (EM) was used in this study simply to highlight the presence of macromolecular formations which corresponded to the high molecular bands seen in the gel separation experiment. IG-TEM for this type of experiments will have to be set with the inclusion of quality controls to overcome these difficulties. The IG-TEM findings on NfH preferential localisation in fluid aggregates may also derive from aspecific binding.

To date, there are no comprehensive electron microscopy studies on circulating protein aggregates and none on macromolecular formations taken from blood stream of ALS patients. The biochemical environment in a fluid state may certainly affect aggregates composition, shape and size, making any TEM analysis very difficult to interpret when compared to any determination performed in tissue (161,162). The NCHs extracted from ALS brains in part had a filamentous and donut-like appearance unlike the more amorphous, globular and at times fibrous appearance of plasma aggregates. While it is clearly impossible to elaborate on these differences and to infer any biological interpretation, it may be possible to speculate on the chemical-physical environment in brain as a potential modifier of aggregation in brain (Figure 4.10). The donut-like particles could be related to the deposition of which is considered a contaminant in brain homogenates as extensively reported in the literature (163–165). Quintana et al. (163) also showed that ferritin binds to tau filaments. As we have not performed any immunolabelling with ALS Brain NCHs, we cannot speculate the relative contribute of Nf to the composition of the aggregates found in brain. We can only speculate that the rough and curved surface of the filaments identified in our TEM analysis of brain aggregates does not match with the smooth and linear shape of intermediate filaments appreciable by TEM (165). However, the micrographs published by Zhou et al. (166) may resemble the small fragments detected in our TEM experiment. In this study, the Authors have used Ficoll 70 at 300 g/l, a polysaccharide utilized to create a crowded cell environment, a very similar condition to the sucrose cushion used in the BPP aggregates enrichment protocol applied in our experiments (our sucrose cushion has a sucrose concentration of 1 M, equal to 342 g/l). It is tempting to speculate on the similarity of our filamentous structures and the micrographs available in literature which portray the main features of prion proteins (165–167). It is also important to add that in our experiments the brain NCHs samples underwent sonication before TEM analysis, a procedure which may have broken the filaments in smaller particles, justifying the absence of long branched (more prion-like) structures (167).

From a biochemical and imaging perspective, to better characterise the macrostructures we have identified in plasma, it will be necessary to assess protein aggregation in neurodegenerative diseases in the affected tissue, such as brain, and compare these formations with those enriched from plasma, thus

extending our preliminary work to a more systematic assays of the conformational and biochemical properties of aggregates in different states in the same pathological context (168–171).

So far, using LC-MS/MS proteomics we have defined the protein composition of aggregate fractions from ALS and HC PPS. The proteins identified in ALS and HC were only partially overlapping and it was possible to identify 103 proteins uniquely in the ALS and 89 uniquely in HC PPS. KEGG pathways functional analysis highlighted a clear distinction in the nature of the enrichment of different pathways in the ALS compared to HC samples. Using stringent definition of enrichment by Webgestalt and a hypergeometric test for statistical evaluation (paragraph 4.3), the proteasome pathway was the only enriched feature in the NCHs from ALS-PPS, while in the HC-PPS NCHs, glycolysis/gluconeogenesis, pentose phosphate pathway and carbon metabolism emerged as consistently enriched pathways. The enrichment of pathways for glucose-dependent ATP production in the healthy state NCHs and their absence in ALS may indicate an imbalance in ATP production already described in ALS (172,173). The presence of a prominent proteasome signal in the enriched KEGG pathways from ALS pooled plasma aggregates is another relevant finding. The change in this pathway involved in the critical process of scavenging of disordered and potentially toxic proteins may be the very reason for the presence of motor neuron protein inclusions in this ALS (174,175). Interestingly, all reported alterations in proteasome activity are so far shown in brain, spinal cord and neuronal cell lines, but not systemically or more specifically, in blood. Our data support the the hypothesis that protein degradation through the proteasome may be critical in ALS and detectable as a systemic response together with a clear propensity of proteins involved in this important pathway to assemble in circulation. To our knowledge, there is no previous description of this phenomenon in the literature.

LC-MS/MS analysis of the NCHs extracted from plasma (HC-PPS and ALS-PPS) showed very little convergence with the analysis performed in brain NCHs. Only a few proteins were shared, and only one (cytoplasmic dynein 1 heavy chain 1, DYNC1H1) was found both in brain and plasma NCHs from ALS individuals. Any interpretation of these findings at this stage may be difficult. It may be possible that brain and plasma-borne heteroaggregates are formed in their respective

environments through processes that are completely independent, but still driven by the aggregogenic properties of some key proteins like neurofilaments that may seed aggregation in different biological contexts. DYNC1H1 mutations have been linked with neurological disorders, such as Charcot-Marie-Tooth, spinal muscular atrophy and mental retardation (176–178), but genetic studies in European populations have not identified a connection with ALS (179). However, a mouse model for motor neuron diseases including mutations for DYNC1H1 and SOD1 seem to connect this protein to ALS, suggesting a role in mithocondria homeostais and interaction with SOD1 toxic species (180). Unfortunately, brain and plasma samples in this analysis have not been screened for DYNC1H1 mutations. These data support the involvement of this protein in biochemical changes occurring during neurodegeneration and the need for further investigations aimed at deepening out understanding of DYNC1H1 as potential biomarker of ALS.

The lack of similarities in the composition of brain and plasma aggregates may also be related to the limits of proteomic techniques, whereby low abundance proteins may be masked by those with high abundance and detection confounded by the presence of post-translational modifications. Therefore, to address these potential shortfalls and gather more information on the nature of brain and circulating aggregates in health and disease, we have undertaken further experiments using a TMTcalibrator™ workflow (discussed in the following chapter), where brain proteins were used to enhance detection of the same proteins in plasma NCHs (93).

A poor overlap was also found comparing the final lists of proteins identified in circulating NCHs, from both ALS and HC, those in brain NCHs and the final list of human/yeast-prion-like proteins produced by An and Harrison (Figure 4.12) (135). Prion-like proteins are supposed to have an intrinsic tendency to aggregation and therefore it would be expected to find these proteins in any aggregate formation, in the healthy or pathological states. It is difficult to explain the lack of protein sequences with prion-like characteristic among NCHs proteins. This may once again be related to the limits of the MS-based proteomics approach employed or to the presence of PTMs in the proteins of interest. Of relevance, cytoskeletal proteins were significantly enriched in both the reported

prion-protein general inventory (135) and in our NCHs protein lists, hinting at the importance of these protein species in the aggregation process. The absence of similarity between the NCHs proteins we have detected and those computed prion-like proteins is testimony of the lack of a full understanding of the mechanisms and factor influencing aggregation and of the translation of this knowledge into computational models that can interpret large datasets.

However, our review of the main features presented by the proteins included in the Brain and plasma (HC and ALS) NCHs by comparison with the human proteome obtained from Uniprot (reviewed sequences only) showed how our protein hits presented features shared by proteins with high tendency to aggregate, in line with what reported by Weids, Ibstedt et al. (136). On average, proteins included in our NCHs had bigger size (Figure 4.13) and a decrease in pI (Figure 4.14), as well as a different trend for hydrophobicity compared to the entire human proteome database of protein-coding genes (Figure 4.15). Interestingly, Brain NCHs proteins were different from those in plasma NCHs in physicochemical properties. Subgroups of plasma NCHs plasma proteins, including those shared and unique to ALS and HC, did not show significant differences among themselves. This would suggest that the physicochemical parameters of the proteins within the NCHs are not affected by the condition (ALS vs HC), but that proteins with specific features (as reported above) tend to become part of protein aggregates, a process which may also depend on the biological environment where these proteins are constitutively expressed or released to. The fact that no significant difference was found between proteins identified in plasma NCHs and the human proteome might be related to the fact that plasma contains the largest variety of proteins (compared to more tissue-specific mixes) resembling the features of the entire human proteome.

In the attempt to characterize the propensity to aggregate of the proteins identified in our NCHs, we have explored another important aspect: the susceptibility to digestion that NCHs proteins have or acquire. Resistance to proteases digestion has been described as a feature of pro-aggregating prion proteins; hence we have tested the hypothesis that in a pathological state like ALS, proteins partaking in plasma and brain NCHs may acquire resistance to proteases digestion (100,144–146). The smear produced after NCHs trypsin

digestion shows an incomplete degradation of the aggregates with this protease (Figure 4.16), while those proteins known to be involved in aggregation, such as NfH and TDP-43 do not show any sign of resistance to digestion (Figure 4.17 and Figure 4.18). Conversely, ubiquitinated proteins in brain NCHs showed a low degree of resistance (Figure 4.19), confirming the role of this PTM in aggregation, perhaps conferring a lower degradation activity of NCHs in ALS which may relate also to their toxicity (156–158,174,175,181). In order to better assess NCHs resistance to digestion and build on our preliminary data, plasma and brain NCHs resistance to protease digestion will be tested using other proteases including chymotrypsin, calpain and enterokinase.

Moreover, in the following chapter we will extend our preliminary analysis to a larger number of samples which will be tested individually and with a technique which allows the simultaneous assessment of brain and plasma. Segmenting our observations to incorporate each individual samples NCHs proteomic profile will allow a better definition of the relation between phenotype (for example speed of disease progression) and systemic molecular features of protein aggregation.

5 Brain derived proteins and their differential regulation in plasma NCHs from ALS and healthy individuals

5.1 Introduction and aim

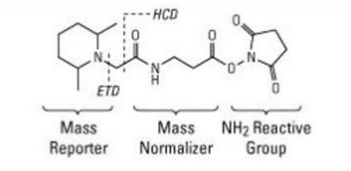
Drawing on to the initial characterization of circulating neurofilament-containing heteroaggregates (NCHs) from Amyotrophic Lateral Sclerosis (ALS) and healthy control (HC) pooled plasma samples (ALS-PPS; HC-PPS), here we apply a novel and more sensitive mass spectrometry (MS) proteomic method to detect the full spectrum of brain-specific proteins in NCHs and to define quantitative differences in the composition of these proteins in the healthy and pathological state.

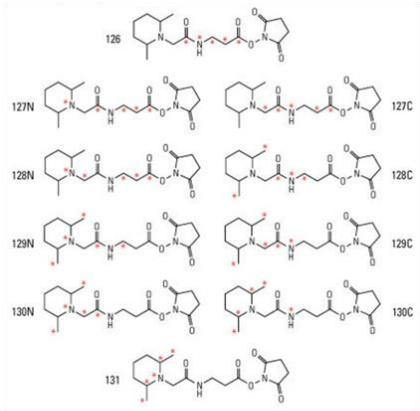
Brain proteins are not expected to be found in high concentration in plasma and they tend to escape detection because of the high dynamic range of plasma proteins, where high-abundance proteins like albumin represent the main bulk of the total protein content in plasma. As previously reported, detection of neurofilament proteins (Nf) in plasma-derived NCHs by mass spectrometry (MS) analysis was not possible. Therefore, to increase signal for detection of low-abundance proteins using MS-based proteomics, an innovative approach, the TMTcalibrator™ workflow, will be used for the purpose of our study (93). This method combines plasma NCHs with brain tissue analysis in the same experiment to enhance detection of tissue-derived proteins in biofluid. This workflow was developed by Proteome Sciences plc (182) to quantify low abundant peptides and proteins in matrices with a high degree of biological complexity (93). TMTcalibrator™ uses Tandem Mass Tag (TMT) (Figure 5.1) label technology to analyse simultaneously in a 10-channel experimental setting six analytical samples and four from a tissue-derived standard called calibrator channels. The latter representing a source of peptides that boosts detection of the same peptides in the analytical samples while functioning as reference for samples quantification. For this purpose, it is important to develop the experiment based on suitable ratio between analytical and calibrator samples which in our case was set at 2X for each analytical sample, along with a ratio of 1:4:6:10 for the four calibrator channels. With this set-up and within each proteomic experiment, the calibrant generated more than half of the total MS-acquired intensity signals. The dominant signal intensity generated by the tissue

derived peptides (calibrator channels) enhanced the detection of the same peptide signals in plasma NCHs, thus ensuring 1) the identification of tissue (brain)-derived peptides in plasma which would normally be below the limit of detection and 2) their quantification using the calibrant internal tissue as the standard reference across different experiments using the same calibrator.

To investigate the presence of brain proteins in plasma NCHs, we designed two identical tenplex experiments: in each tenplex, we used Netherlands Brain Biobank (BNN) post-mortem brain tissue from ALS patients for the calibrator channels and plasma NCHs from three ALS patients and three HC. Based on our previous experiment results, outlined in Chapter 4, and on the only marginal overlap between brain and plasma NCHs protein profiles (Paragraph 4.4.3), we have opted for whole brain lysate rather than brain NCHs as tissue calibrator. This was done to increase the detection of brain-specific proteins in plasma over and above the 48 identified in the MS proteomic experiment of brain NCHs (Table 4.4).

(A)





(B)

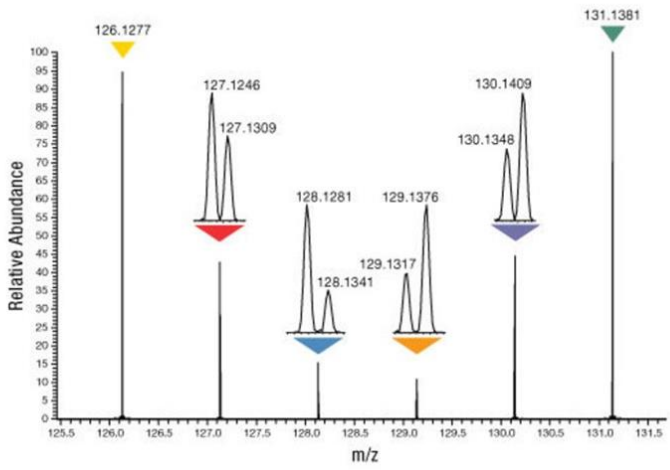


Figure 5.1. Tandem Mass Tag (TMT) 10 reagents for peptide labelling (ThermoFisher Scientific).

(A) TMT general chemical structure showing the mass reporter, mass normaliser and NH_2 reactive group regions. The tags are shown with their masses and isotope positions (red dots). (B) Relative quantification of the samples after detection of the mass reporter ions generated in the second mass spectrometry (MS) step in a liquid chromatography coupled with tandem MS (LC-MS/MS).

5.2 Samples selection

ALS is an incurable and invariably fatal neurodegenerative disease were a significant diagnostic delay has hampered any progress towards developing effective treatment. This is mostly due to the extreme phenotypic heterogeneity of the disease and the lack of biomarkers to be used as surrogate markers of treatment response and for clinical stratification. Therefore, in order to limit the potential confounding effects of clinical heterogeneity of the disease, six ALS patients with the same El-Escorial criteria for diagnosis (88) (stage of disease) were selected, all of whom showed limb-onset disease. Moreover, because ALS occurrence is also strongly age-related (183,184), an homogeneous group in terms of age of ALS onset was chosen (average63.87 ± 3.09 years old) (Table 5.1). Age and gender-matched healthy controls were selected (HC; average 64.16 ± 3.05 years old Table 5.1). The selected ALS patients showed also moderately homogeneous, moderate to severe, level of neurological impairment at the time of sampling as shown by the ALS Functional Rating Scale Revised (ALFRS-R, 28 ± 10.58), while speed of progression at the time of sampling (0.946 ± 0.47) was in keeping with a moderately to fast progressing ALS (Table 1).

NCHs enrichment and separation from plasma samples was performed using the Brain-Plasma Protocol (BPP) described in Paragraph 2.2.2.1, the methodology shown to give the most comprehensive cover of the plasma protein profile. Total protein concentration was measured in plasma samples before and after the enrichment, in order to evaluate the relationship between plasma and NCHs protein concentration. Protein concentration in plasma NCHs and the ration between NCHs and plasma protein for each sample showed a higher variability in HC compared to ALS (Figure 5.2). There were no statistically significant differences between ALS and HC protein concentration in plasma and NCHs and in the respective ratios (Figure 5.2).

The plasma NCHs from the selected ALS patients and HC to be loaded on to the analytical channels of the TMTcalibrator™ as well as a 1:1 mixture of two ALS brain samples (processed as reported in Paragraph 2.5.2) to be loaded in the calibrant channels were subjected to quality check by gel electrophoresis (Figure 5.3).

Sample ID	Gender	Ethnicity	Age at sampling (year)	El-Escorial	Clinical Onset	ALSFRS-R	Rate of progression at sampling
ALS1	F	Caucasian	61.37	Definite ALS	LL	32	0.737
ALS2	F	Caucasian	68.80	Definite ALS	UL	35	0.821
ALS3	M	Caucasian	60.18	Definite ALS	LL	35	0.729
ALS4	M	Caucasian	65.19	Definite ALS	UL	9	1.912
ALS5	M	Caucasian	64.79	Definite ALS	LL	22	0.750
ALS6	F	Caucasian	62.89	Definite ALS	LL	35	0.727
HC1	F	Caucasian	62.48	NA	NA	NA	NA
HC2	F	Caucasian	67.50	NA	NA	NA	NA
HC3	M	Caucasian	60.64	NA	NA	NA	NA
HC4	M	Caucasian	68.28	NA	NA	NA	NA

Sample ID	Gender	Ethnicity	Age at sampling (year)	El-Escorial	Clinical Onset	ALSFRS-R	Rate of progression at sampling
HC5	M	Caucasian	62.54	NA	NA	NA	NA
HC6	F	Caucasian	63.53	NA	NA	NA	NA

Table 5.1. Clinical and demographic features of ALS patients and healthy controls from whom blood samples were collected for neurofilament-containing heteroaggregates (NCHs) enrichment and TMTcalibrator™ analysis.

Light orange colour-code: amyotrophic lateral sclerosis (ALS) patients

Light blue colour-code: healthy control (HC) samples

NA: no data available.

Sample ID: ALS= ALS patients; HC= Healthy Control

Gender: M= Male; F= Female

El-Escorial: diagnostic classification of ALS (88)

Clinical Onset (site of initial clinical signs): LL= Lower Limbs; UL= Upper Limbs

Age at sampling: age of the patient when blood sample was obtained

ALSFRS-R (ALS Functional Rating Scale revised): level of neurological impairment across different clinical domains (1-48, higher neurological impairment with lower values)

Rate of progression at sampling: calculated as 48 - ALSFRS-R score at sampling time/disease duration from onset of symptoms to sampling time

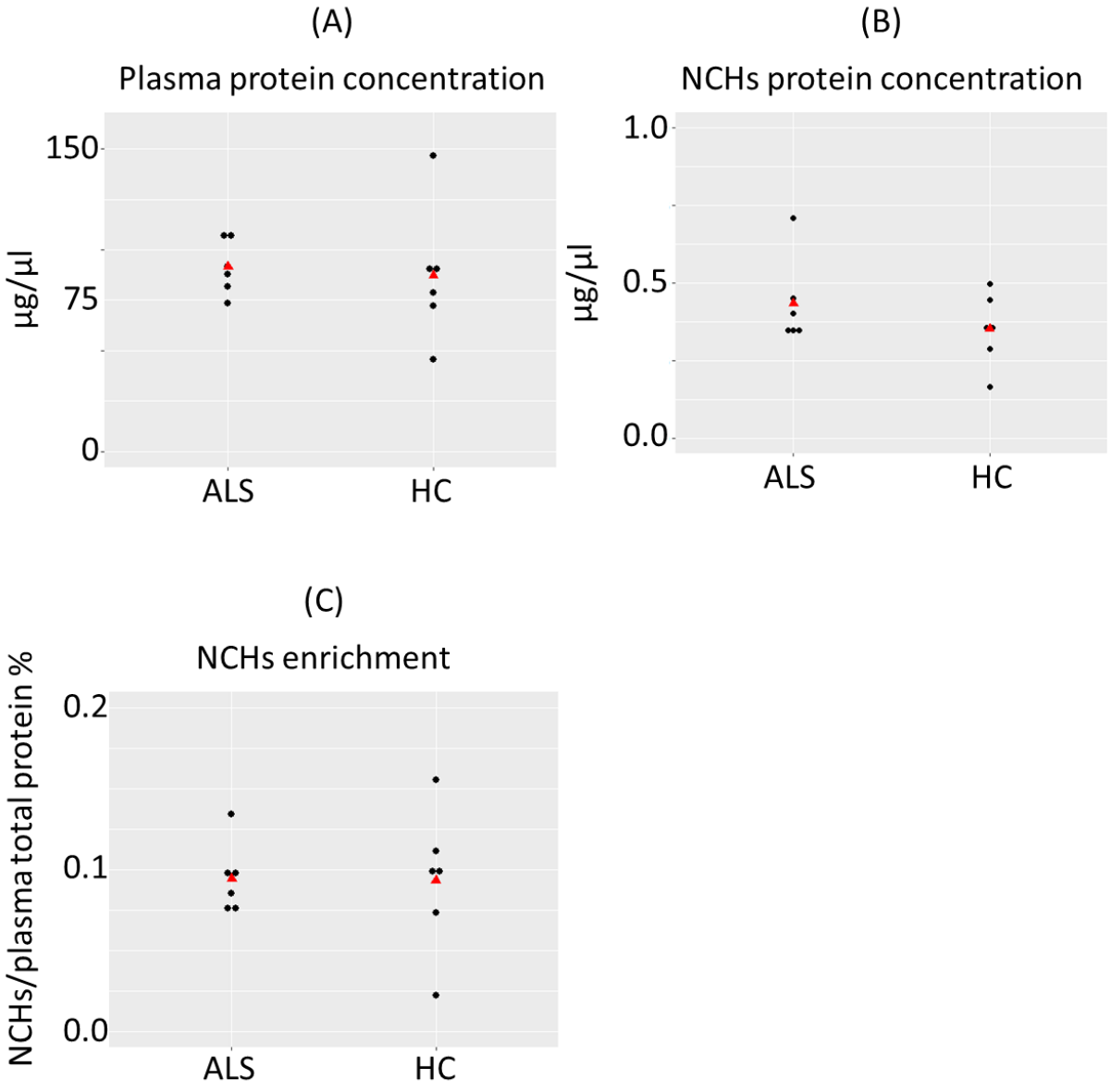


Figure 5.2. NCHs enrichment performance in the ALS and HC plasma samples.

Dot plots showing the protein concentration distribution of the plasma (A) samples before enrichment and NCHs (B) samples after enrichment. In (C) the amount of total protein in NCHs expressed as percentage (%) of the total protein in related plasma samples. The black dots represent the single data points of the samples obtained from the patients listed in Table 5.1, while the red triangles represent the mean value for each group (ALS: 91.4 µg/µl (A), 0.43 µg/µl (B), 0.0945 % (C); HC: 87.3 µg/µl (A), 0.35 µg/µl (B), 0.0932 % (C).

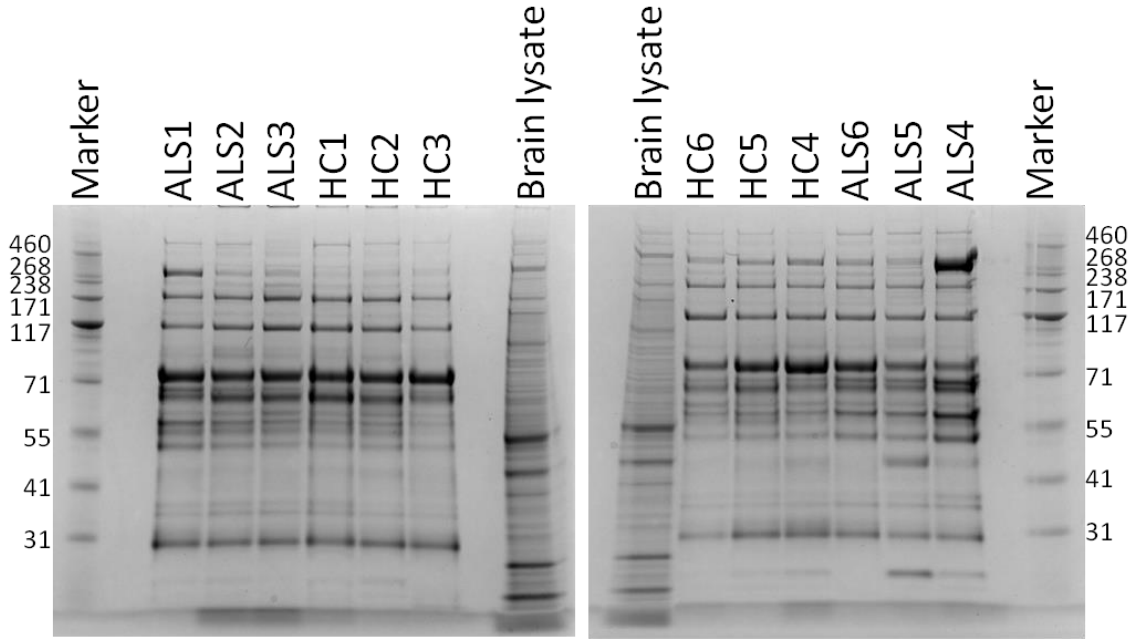


Figure 5.3. NCH samples quality check by Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis (SDS-PAGE).

Plasma NCHs (ALS1 - 6 and HC1 - 6) and Brain lysates were loaded into a 3-8% Tris-acetate gel and stained with Imperial Blue Comassie after gel electrophoresis to evalute sample quality with regards to 1) protein expression profiles, solubilisation and presence of contaminants in each sample. A good and reproducible pattern of protein separation was seen in all samples with no evidence of contaminants. Marker: protein marker.

5.3 TMTcalibrator™: experimental setting

Once samples were quality checked, they were prepared for further analysis as described in Paragraph 2.5.2.1. Briefly, each sample was reduced with Dithiothreitol (DTT), alkylated with iodoacetamide (IAA) and later digested with trypsin to generate tryptic peptides which were desalted in a chromatographic step. At this stage, the peptides generated from the Brain sample were split in aliquots of 1:4:6:10 volumes, to provide the four calibrant samples to be used in the TMTcalibrator™ analysis. Finally, all samples underwent TMT labelling as reported in Table 5.2.

After labelling, samples ALS1, ALS2, ALS3, HC1, HC2, HC3 and four calibrant samples (Cal1X, Cal4X, Cal6X and Cal10X) were joined in the 10plex1, while samples ALS4, ALS5, ALS6, HC4, HC5, HC6 and the same four calibrant samples in the 10plex2 (Table 5.2). The two 10plexes underwent basic reverse phase (bRP) fractionation generating eight fractions for each 10plex.

Sample ID	Protein amount (µg)	10plex	TMT reagent
ALS1	40	1	TMT10-126
ALS2	40	1	TMT10-127N
ALS3	40	1	TMT10-127C
HC1	40	1	TMT10-128N
HC2	40	1	TMT10-128C
HC3	40	1	TMT10-129N
Cal1X	20	1	TMT10-129C
Cal4X	80	1	TMT10-130N
Cal6X	120	1	TMT10-130C
Cal10X	200	1	TMT10-131
ALS4	40	2	TMT10-126
ALS5	40	2	TMT10-127N
ALS6	40	2	TMT10-127C
HC4	40	2	TMT10-128N
HC5	40	2	TMT10-128C
HC6	40	2	TMT10-129N
Cal1X	20	2	TMT10-129C
Cal4X	80	2	TMT10-130N
Cal6X	120	2	TMT10-130C
Cal10X	200	2	TMT10-131

Table 5.2. ALS and HC analytical and brain calibrant channels in the two 10plexes experiments.

In this table, we report 1) the amount of total proteins labelled for the TMTcalibrator experiment in each sample (Protein amount), 2) the 10plex experiment for each sample and 3) the relative TMT tag (TMT reagent).

5.3.1 Injection conditions

The liquid chromatography (LC) coupled with tandem mass spectrometry (MS/MS) analysis was set with a five hours elution gradient for the LC step, to allow a good resolution of the eluted peptides, and coupled with a Thermo Scientific™ Orbitrap Fusion Tribrid (Thermo Scientific) mass spectrometer. This combination provided higher resolution at the second MS scan improving

peptide-spectrum matches (PSMs) and relative quantification with the TMT labels.

For the LC-MS/MS, 12 µg were loaded into the chromatographic column and each of the eight fractions obtained after bRP from each of the two 10plexes was injected in a double shot (e.g. injected twice), generating a total of 32 raw files. Peptides and proteins identification was carried out with the Proteome Discoverer 1.4 (PD) and quantitative analysis was performed using the TMTcalibrator™ bioinformatics pipeline, consisting of the Calibrator Data Integration Tool (CalDIT) and the Feature Selection Tool (FeaST) (Figure 5.4). Following, the functional Analysis Tool (FAT) provided a comprehensive characterization of the output data.

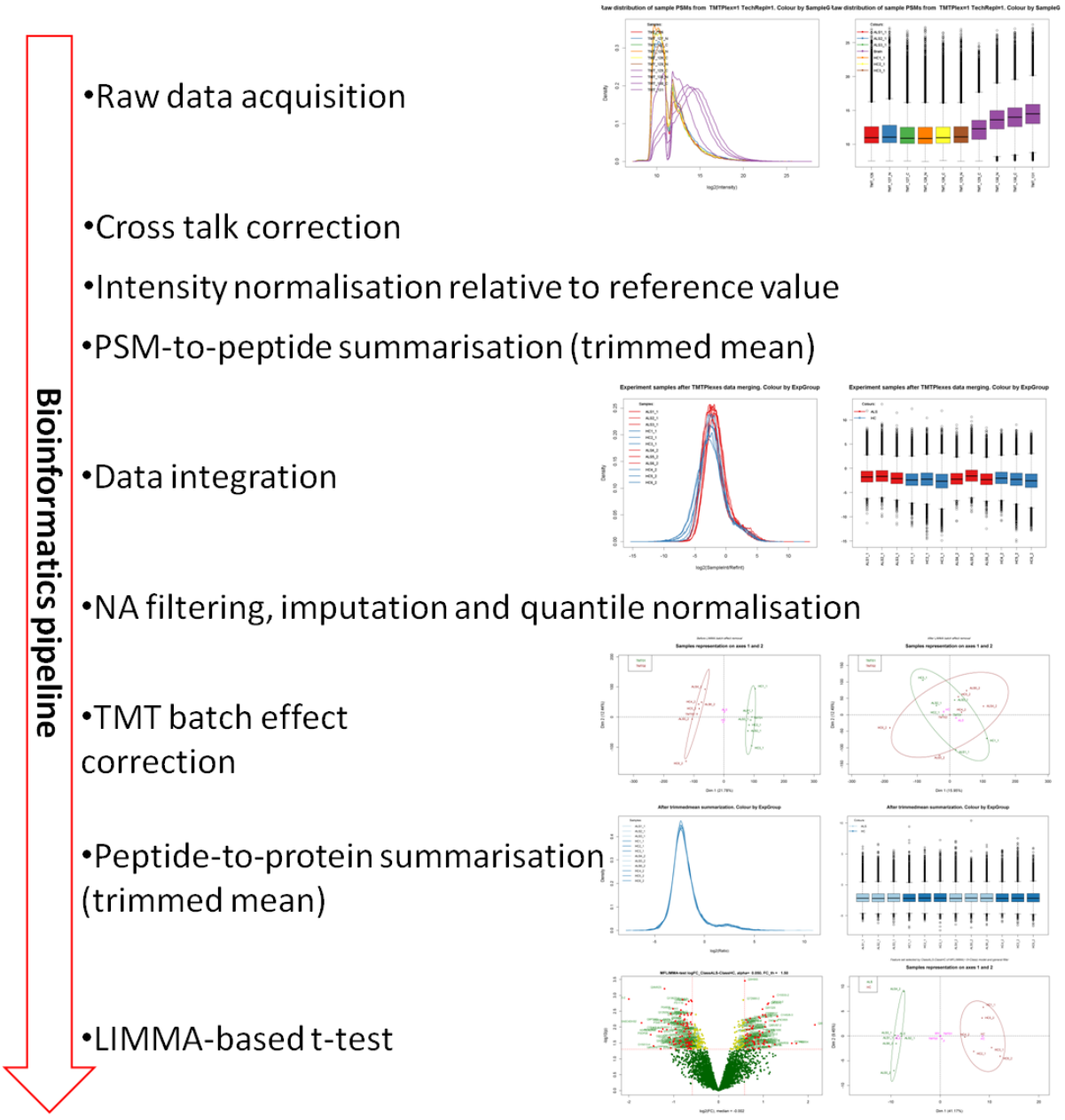


Figure 5.4. TMTcalibrator™: bioinformatic pipeline.

After MS/MS spectra (raw data) acquisition, the intensity of each channel was corrected for background and cross-talking between tags in the second mass spectrometer (MS2). Intensity values of the detected Peptide-Spectrum Matches (PSMs) were normalised with a reference value generated as the average of the Calibrant channels and this was followed by PSM-to-peptide summarisation defined as “trimmed mean”. In this passage the data points considered as outliers in each analytical sample were removed stabilizing the mean before merging the data obtained from the two 10plexes analysed. Then, “not available data points (NA)” filtering, imputation and quantile normalization were performed on the merged data set, so that it was possible to perform a Principal Component Analysis (PCA) on the data acquired. It was also possible to evaluate the TMT batch effect within linear models for microarray data (LIMMA). After peptide-to-protein summarisation, statistical significance of the protein groups identified was

analysed with LIMMA-based t-test (p-Value < 0.05, Fold Change threshold = 1.5). PCA showed a separation between the ALS and HC protein groups in the first component of 41.17%.

5.4 Bioinformatics pipeline

After peptides and proteins identification obtained using Proteome Discoverer v1.4 (PD), a total of 679471 PSMs were identified, of which 323442 belonged to 10plex1 (Table 5.3) and 356029 to 10plex2 (Table 5.4). These PSMs were grouped by PD in 99749 peptides. With CalDIT, the intensity distribution in the two 10plexes was corrected for cross-talk between different tags (185). The four calibrant channels were then normalised using the median values to evaluate variability and to generate reference values to be used in the normalisation process of the analytical sample intensities (expressed as $\log_2$; PSM intensity in the sample/PSM intensity of the reference value) (Figure 5.4).

Following normalisation, the data obtained from the analytical channels were subjected to PSM-to-peptide summarisation, providing data sets with quantitative information on 82565 PSMs in 10plex1 and 85567 PSMs in 10plex2. At this stage, the calibrator channels were corrected according to the median intensity and normalised across the two 10plexes, to generate a data matrix with the intensities from each analytical sample for each peptide reported in the peptide list.

	ALS1	ALS2	ALS3	HC1	HC2	HC3	Cal1X	Cal4X	Cal6X	Cal10X
Total PSMs	323442	323442	323442	323442	323442	323442	323442	323442	323442	323442
PSMs with quantitative data	131473	144640	130907	110389	141858	178085	236983	291190	299265	307340
Peptides	49688	52877	49597	42822	52134	58864	70059	78892	80159	81178
Protein groups	7515	7645	7510	7080	7529	7808	8318	8907	9003	9096

Table 5.3. Peptide and protein identification performance in 10plex1.

Number of peptide-spectrum matches (PSMs), peptides and protein groups identified for each sample in 10plex1 using Proteome Discoverer 1.4 (PD). ALS and HC samples are shown in light-orange and light-blue colour-code respectively.

Total PSMs: number of PSMs detected in 10plex1

PSMs with quantitative data: number of peptide-spectrum matches (PSMs) containing quantitative data for the specific sample

Peptides: number of identified peptides for that specific sample

Protein groups: number of identified protein groups for that specific sample

	ALS4	ALS5	ALS6	HC4	HC5	HC6	Cal1X	Cal4X	Cal6X	Cal10X
Total PSMs	356029	356029	356029	356029	356029	356029	356029	356029	356029	356029
PSMs with quantitative data	112219	166427	132753	153268	158746	198459	253218	316195	326344	336944
Peptides	43203	57998	49641	54729	56251	63677	73704	82383	83563	84530
Protein groups	7168	7945	7575	7808	7809	8097	8581	9192	9291	9393

Table 5.4. Peptide and protein identification performance in 10plex2.

Number of peptide-spectrum matches (PSMs), peptides and protein groups identified for each sample in 10plex2 using Proteome Discoverer 1.4 (PD). ALS and HC samples are shown in light-orange and light-blue colour-code respectively.

Total PSMs: number of all the PSMs detected in 10plex2

PSMs with quantitative data: number of peptide-spectrum matches (PSMs) containing quantitative data for the specific sample

Peptides: number of identified peptides for that specific sample

Peptides: number of identified peptides for that specific sample

Protein groups: number of identified protein groups for that specific sample

Feature Selection Tool (FeaST) was then used to filter out the peptides with more than two “not available data point” (NA) from the ALS and HC experimental groups. FeaST was also used to impute the NAs for those peptides identified in at least four out of six analytical channels in each experimental group. In this way, FeaST generated a matrix with quantitative data for 37638 peptides represented in both ALS and HC experimental groups. This subset was then normalised again on calibrant quantile distribution and examined by linear models for microarray data (LIMMA)-based modified t-Test to evaluate a possible batch-effect of the TMT labels in the two 10plexes which may have affected the quantification of the samples. The TMT batch-effect was evaluated by principal component analysis (PCA) and quantified for the 21.78% of the variance between the two experimental groups (Figure 5.5A). After batch-effect removal, samples could not be separated based on the two different 10plexes and variation was reduced to 15.95% (Figure 5.5B).

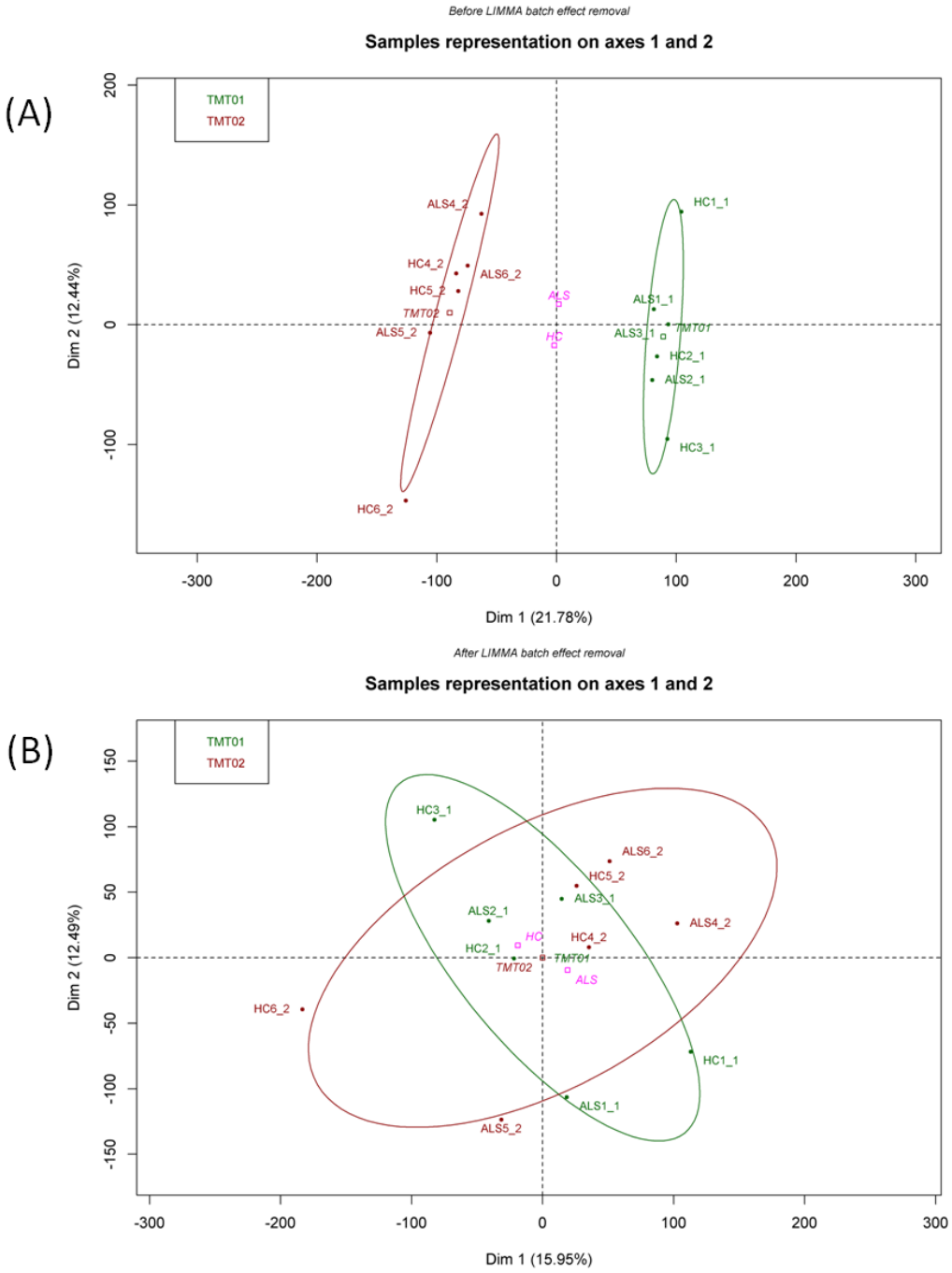


Figure 5.5. Principal component analysis (PCA) of the ALS and HC proteomic data for Tandem Mass Tag (TMT) batch-effect.

(A) The figure shows the TMT batch-effect on the ALS and HC proteomic data without LIMMA test correction. PCA was performed to study the variance between the two 10plexes (TMT01 and TMT02) as dimension 1 and between the experimental groups (ALS and HC) as the dimension 2. ALS and HC samples appeared to group according to the 10plex in which they were analysed, suggesting a strong batch effect. (B) After the TMT batch-effect removal by linear models for microarray data LIMMA, the PCA showed that the samples analysed by TMTcalibrator™ did not group according to the different TMT 10plexes and that there was minimal separation between ALS and HC datasets.

FeaST protein grouping and peptide-to-protein summarisation (Figure 5.4) generated a list of 4973 protein groups which were further analysed by LIMMA-based t-test at protein level. Statistical analysis of the data at peptide and protein level resulted in 229 peptides (0.61%, of the identified peptides) and 285 protein groups (5.7%, of the identified proteins) which were statistically significant (Figure 5.6).

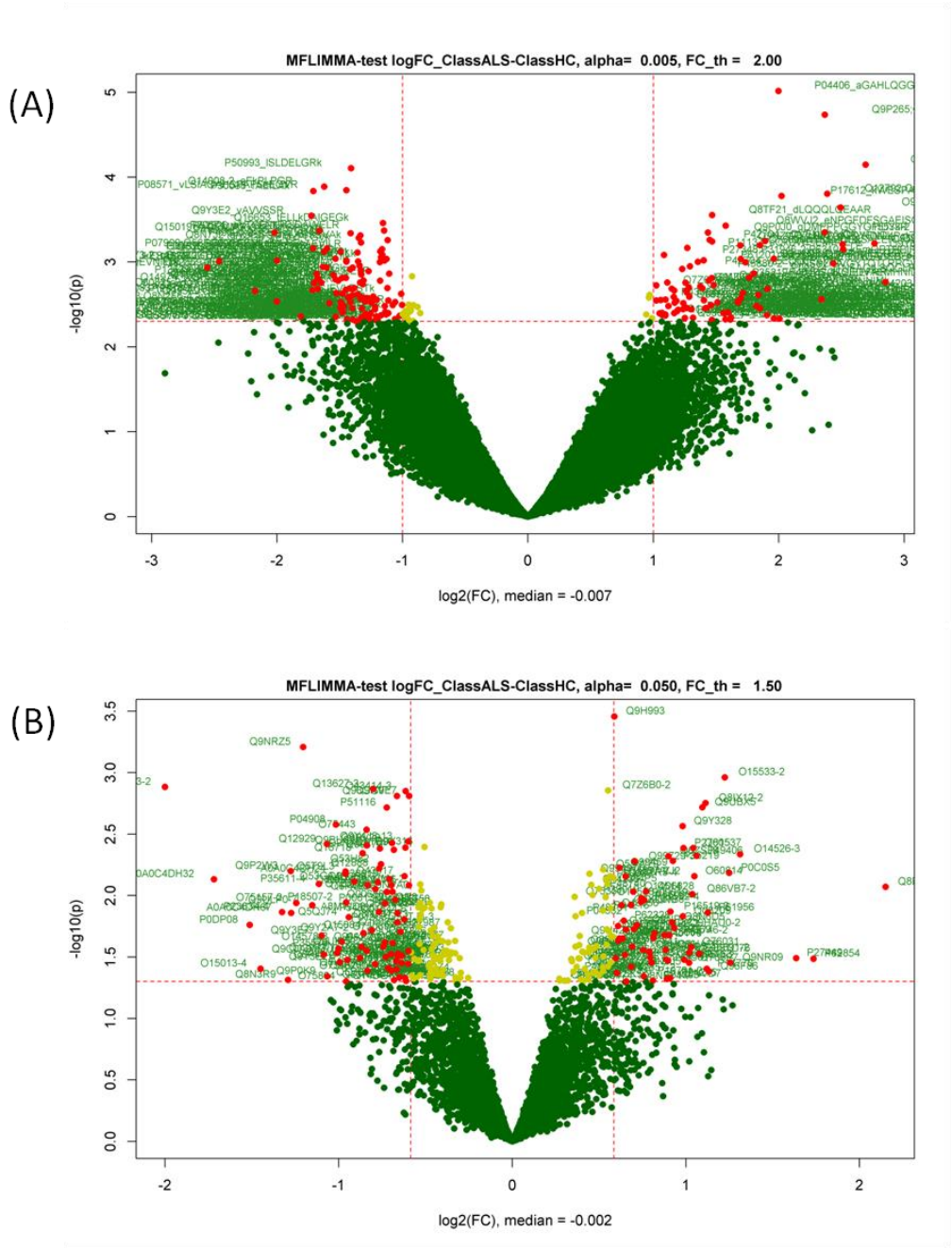


Figure 5.6. Volcano plots of the regulated features detected using TMTcalibrator™ and analysed by LIMMA.

The volcano plots show the distribution of the peptides (A) and proteins (B) identified in the TMT proteomic study according to their fold change (FC) expressed as log_2 (fold change ALS/HC) (logFC) in the x axis and according to p-value expressed as –log10 (p-value) in the y axis. In (A), peptides were considered regulated if p-value < 0.005 and logFC <-1 or > 1. In (B), protein groups were considered regulated if p-value < 0.05 and logFC < -0.58 or > 0.58. Red dots are regulated features, yellow dots are features with a significant p-value (p < 0.05) and logFC between -0.58 and 0.58 while green dots are not significant features. Uniprot Ids are reported beside the dots, with the specific peptide sequence in (A), with significant p-value.

PCA analysis of the regulated features in the ALS vs HC TMT proteomic experiment showed a variance of 59.92% at peptide level and of 41.17% at protein level in the first dimension (Figure 5.7), supporting the observation of a significant difference in brain derived proteins in plasma NCHs from HC and from ALS individuals. ALS patients and HC proteomic datasets appeared grouped separately when the same regulated features were tested using relative clustering after LIMMA t-test analysis and represented using heatmaps (Figure 5.8).

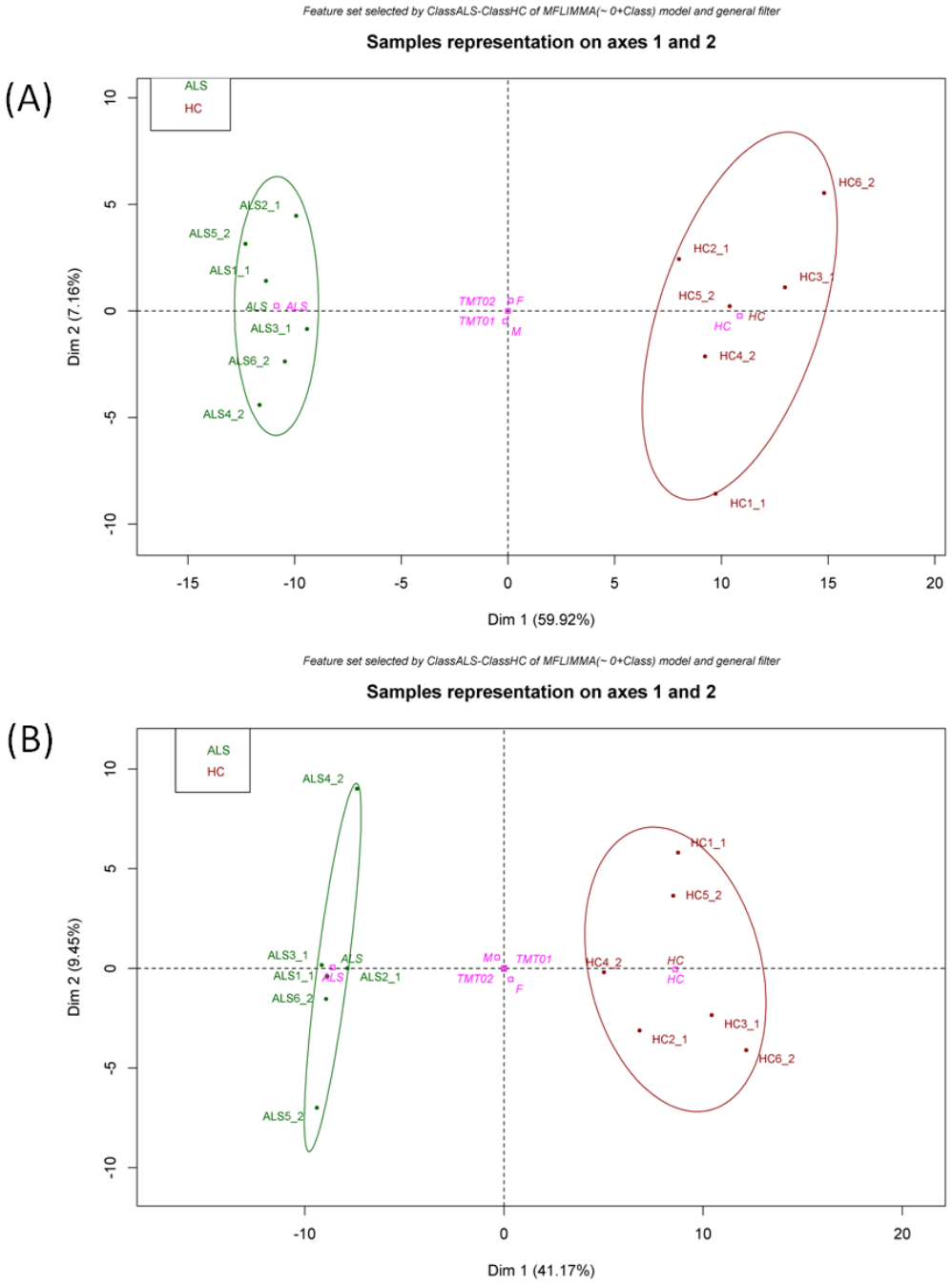


Figure 5.7. PCA of the regulated features identified in the ALS vs HC NCHs TMT proteomic analysis.

PCA shows a good separation between the ALS and HC experimental groups regulated features at peptide (A) and protein (B) levels. The two PCAs were performed considering the variance between the two experimental groups (ALS and HC) as dimension 1 and the variance between 10plexes (TMT01 and TMT02) as the dimension 2. The ALS and HC samples analysed are clearly grouped and separated in distinct experimental groups.

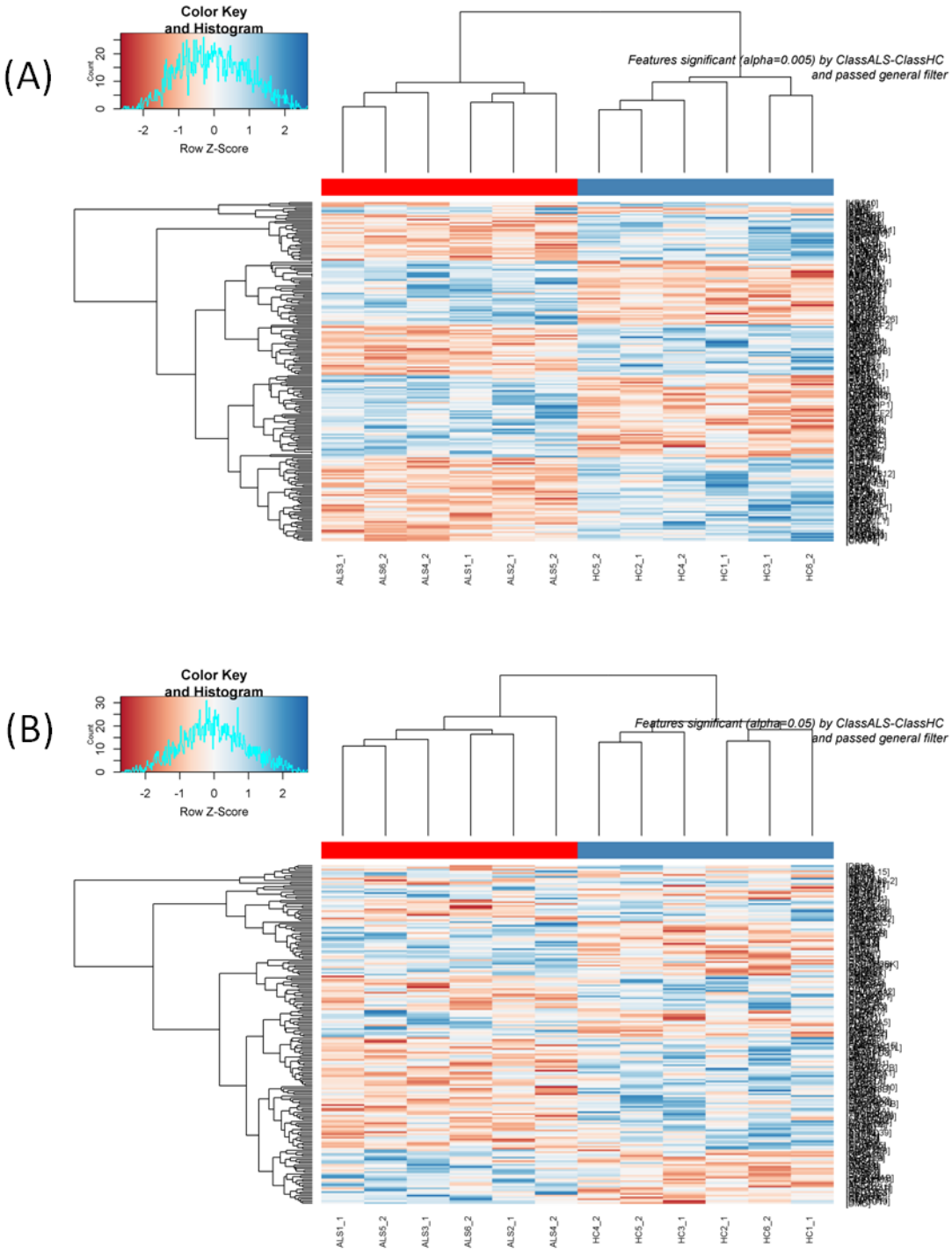


Figure 5.8. Heatmaps and clustering of the regulated features identified in the ALS vs HC NCHs TMT proteomics analysis detected using the LIMMA test.

The heatmaps in (A) and (B) show the distribution of the regulated features, and relative clustering, at peptide and protein level respectively. The regulated features are distributed vertically, with the Uniprot IDs on the right-hand side and relative clustering on the left-hand side, while the analytical samples are distributed horizontally, with sample names at the bottom and relative clustering at the top of the heatmaps. ALS patients and the HC distributed in separate

clusters in both diagrams. The colour key histogram at the top left side shows the distribution of the features and the heatmap colour coding.

5.5 FeaST analysis of regulated features at protein level

The final list of proteins identified in plasma NCHs and related quantitative data produced by FeaST contained almost 5000 brain-derived proteins, including neurofilaments (Nf). All three Nf protein isoforms of interest were identified. Nf Light (NfL), Nf Medium (NfM) and Nf Heavy (NfH) were found at a relatively higher level in the ALS samples compare to HC (log_2-fold change ALS/HC) (logFC) = 0.093, 0.181 and 0.298, respectively) but none was significantly regulated (p= 0.40, 0.16 and 0.06 respectively). Matching the FeaST protein list with the ALS gene list obtained from www.malacards.org (186), an integrated database of human diseases and their annotations, we found 24 proteins matching our list of protein and only Fused in Sarcoma RNA-binding protein (FUS) was found to be significantly regulated in ALS vs Control (p=6.96E-03) (Table 5.5).

Gene name	Uniprot ID	Protein name	Unique peptides	logFC	p-value
FUS	P35637-2	Isoform Short of RNA-binding protein FUS	5	0.564	6.96E-03
NEFH	P12036	Neurofilament heavy polypeptide	27	0.298	6.36E-02
OPTN	Q96CV9	Optineurin	11	-0.243	9.58E-02
UNC13A	Q9UPW8	Protein unc-13 homolog A	11	-0.326	9.75E-02
PON2	Q15165-3	Isoform 3 of Serum paraoxonase/arylesterase 2	1	0.292	1.46E-01
ANG	P03950	Angiogenin	1	-0.595	1.67E-01

Gene name	Uniprot ID	Protein name	Unique peptides	logFC	p-value
CHMP2B	Q9UQN3	Charged multivesicular body protein 2b	1	0.492	1.90E-01
VCP	P55072	Transitional endoplasmic reticulum ATPase	65	-0.514	2.16E-01
ATXN2	Q99700-2	Isoform 2 of Ataxin-2	2	-0.445	2.51E-01
ANXA11	P50995-2	Isoform 2 of Annexin A11	16	0.150	3.01E-01
SOD1	P00441	Superoxide dismutase [Cu-Zn]	8	0.176	3.39E-01
ERBB4	Q15303-4	Isoform JM-B CYT-2 of Receptor tyrosine-protein kinase erbB-4	1	-0.284	3.70E-01
TARDBP	Q13148	TAR DNA-binding protein 43	2	0.180	3.85E-01
SQSTM1	Q13501	Sequestosome-1	2	-0.280	4.00E-01
MATR3	P43243	Matrin-3	16	0.127	4.03E-01
PFN1	P07737	Profilin-1	14	0.136	4.08E-01
VAPB	O95292	Vesicle-associated membrane protein-associated protein B/C	10	-0.094	4.31E-01
EPHA4	P54764	Ephrin type-A receptor 4	13	-0.082	5.37E-01

Gene name	Uniprot ID	Protein name	Unique peptides	logFC	p-value
PON1	P27169	Serum paraoxonase/arylesterase 1	14	0.123	6.07E-01
TAF15	Q92804-2	Isoform Short of TATA-binding protein-associated factor 2N	3	0.067	6.99E-01
UBQLN2	Q9UHD9	Ubiquilin-2	4	-0.065	6.99E-01
HNRNPA1	P09651-3	Isoform 2 of Heterogeneous nuclear ribonucleoprotein A1	8	-0.066	7.50E-01
DCTN1	Q14203-6	Isoform 6 of Dynactin subunit 1	2	0.020	9.34E-01
TBK1	Q9UHD2	Serine/threonine-protein kinase TBK1	5	-0.004	9.86E-01

Table 5.5. List of proteins from ALS genes found in the ALS vs HC TMTcalibrator™ data and level of regulation.

The table outlines the ALS genes proteins and their regulation identified matching the ALS vs HC TMT-proteomic dataset against the gene classifiers include in the MalaCards Human Disease Databse (**https://www.malacards.org/card/amyotrophic_lateral_sclerosis_1#RelatedGenes-table)** (186), selecting only the “elite genes” which are those more likely to cause the disease. 38 elite genes were identified, of which 24 linked to ALS were detected in the list generated by the TMTcalibrator™ experiments. The “elite genes” not identified in the TMTcalibrator™ data set: PPARG Coactivator 1 Alpha (PPARGC1A), Triggering Receptor Expressed On Myeloid Cells 2 (TREM2), NIMA Related Kinase 1 (NEK1), Amyotrophic Lateral Sclerosis 3 and 7 (ALS3, ALS7), Paraoxonase 3 (PON3), Chromosome 21 Open Reading Frame 2 (C21orf2), RNA Export Mediator GLE1 (GLE1), Cyclin F (CCNF), Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 10 (CHCHD10), D-Amino Acid Oxidase (DAO), FIG4 Phosphoinositide 5-Phosphatase (FIG4), Chromosome 9 Open Reading Frame 72 (C9orf72) and Peripherin (PRPH).

Uniprot ID: Uniprot database protein identifier

Gene name: the gene symbol used to represent a gene
Protein name: protein full name recommended by Uniprot
Unique peptides: number of peptide sequences unique to a protein group
logFC: relative quantification with value expressed as log_2(ALS/HC) intensities
p-value: statistical significance for differential regulation between ALS and HC experimental groups

Of the 5000 brain-derived proteins identified in the TMT Calibrator experiment, 285 proteins showed a statistically significant regulation ($p < 0.05$). 158 were more highly expressed in HC (logFC < 0) with an average fold change of -0.667, while 127 showed a higher expression in ALS (logFC > 0) with an average fold change of 0.703. It is reasonable to think that all these regulated proteins are potential biomarker candidates of ALS as they are brain proteins in circulation, found in systemic aggregates, which present a differential expression in relation to the pathological state. Future work will have to include the selection of the best candidates to be considered for further validation in larger cohorts of ALS and HC, using a different methodological approach like immunodetection. The selection of the best candidates for future studies will hinge not only on the reported level of disease-regulation, but also on the functional role of the protein in the disease pathogenesis and availability of screening methods that would make high-throughput experiments easy to perform. Based on the overall findings so far, the ideal candidate biomarker of ALS would have to be central to the main pathogenic alterations in brain tissue and plasma that our proteomic study has identified like for example biochemical changes. The neurochemical marker will ideally be readily measurable and could possibly be used as measure of treatment response and of target engagement in new therapeutic strategies (187–190). For this reason, we will also include a functional study of the proteins identified within the TMT Calibrator proteomic data set, to gain insight into the biochemical and physiological changes that these candidates underpin in the context of the ALS pathology.

5.6 Functional analysis

As described in Paragraph 2.5.2.2, Functional Analysis Tool (FAT) was used to evaluate enriched and regulated functional categories among the reported peptides and proteins produced by FeaST. As our MS-based proteomic study focused primarily on peptide sequencing and identification, we deemed the

information obtained at peptide level more accurate. However, considering the overall aim to identify readily detectable biomarkers in plasma to help in ALS diagnosis, prognosis and patient stratification only information obtained at protein level were further considered.

Briefly, the list of proteins generated was searched to identify specific functional categories, such as gene ontology (GO) terms and pathways through searches in multiple databases, and the output at defined statistical thresholds was a list of enriched and regulated pathways.

FAT analysis of “regulated” features (Paragraph 2.5.2.2) proved to be more sensible to subtle changes in the data set compared to FAT analysis of “enriched” data and also more capable of highlighting the presence of categories otherwise not detectable by the enrichment analysis. As already reported in Paragraph 2.5.2.2, the ability to depict different functional terms in FAT analysis rests on the statistical methods applied to mine different databases with the submitted protein lists. In contrast, FAT enrichment analysis performed at protein level was not effective in disclosing meaningful changes in our datasets. For this reason, FAT regulated analysis was used to evaluate the physiological and biochemical changes intrinsic to our datasets.

Regulated FAT analysis (p-value < 0.05) identified 69 biochemical pathways, 77 gene ontology (GO) biological process (BP), 10 GO molecular function (MF), 19 GO cellular component (CC), 1 protein motif, 4 protein domain and 12 miRNA targets. The regulated pathway analysis pointed to relevant molecular mechanisms involved in ALS pathology and covered proteins acting as potential biomarkers.

Within the top10 regulated pathways (Table 5.6), five out of the top10 were involved in metabolism of lipoproteins (lipoprotein metabolism, plasma lipoprotein assembly, remodelling and clearance, plasma lipoprotein assembly, lipid digestion, mobilization and transport, plasma lipoprotein remodelling), all showing a higher level of expression in ALS (median FC > 0).

The significant representation of these pathways in our brain/NCHs proteomes reflects known changes of the ALS pathophysiology, where metabolism is thought to be switched from sugar and carbohydrate to lipid use (191–193).

Based on this observation, it may be reasonable to consider NCHs as a good readout of the pathological changes that are central to the pathogenesis of ALS. However, none of the single proteins belonging to these pathways showed a significant regulation (p-value < 0.05), making them less relevant as potential target biomarkers in ALS.

Pathway description	Number of genes in the category	p-Value	Median logFC
Lipoprotein metabolism [Reactome]	27	4.11E-04	0.19
Dopamine Neurotransmitter Release Cycle [Reactome]	15	4.53E-04	0.24
DCC mediated attractive signaling [Reactome]	8	6.65E-04	0.36
Plasma lipoprotein assembly, remodeling, and clearance [Reactome]	25	8.03E-04	0.19
Plasma lipoprotein assembly [Reactome]	13	1.65E-03	0.29
Netrin-1 signaling [Reactome]	15	4.80E-03	0.20
Lipid digestion, mobilization, and transport [Reactome]	36	5.25E-03	0.11
Oncostatin M Signaling Pathway [Wikipathways]	21	8.09E-03	-0.13
Plasma lipoprotein remodeling [Reactome]	9	8.16E-03	0.19
CD28 dependent Vav1 pathway [Reactome]	6	8.35E-03	0.28

Table 5.6. FAT analysis of the ALS vs HC TMT proteomic analysis: top10 regulated pathways.

Pathway description: name of the pathway identified (reference database in square brackets, e.g. [Reactome])

Number of genes in the category: Number of genes included in the given pathway in Homo sapiens

p-Value: Statistical significance calculated by the Mann-Whitney U test as described in Paragraph 2.5.2.2

Median logFC: median value of expression (logFC values of the proteins included in the given pathway in Homo sapiens)

The dopamine neurotransmitter release cycle pathway relates to biochemical mechanisms linked to dopamine (DA) biosynthesis: dopamine is loaded into vesicles which are then docked, primed and fused with the plasma membrane of synapses to release dopamine into the synaptic cleft. In the literature, IgG from ALS patients were found to block DA secretion from PC12 cell line (194), pointing towards the possible involvement of this pathway in ALS. Data analysis showed that protein lin-7 homolog B (LIN7B; unique peptide= 1, logFC= 0.529, pValue= 0.0277) was the only significantly regulated protein. This protein takes part in the tripartite complex composed of LIN7 (LIN7A, LIN7B or LIN7C), CASK and APBA1, and may have the potential to couple synaptic vesicle exocytosis to cell adhesion in brain. Unfortunately, there is no literature linking these proteins to ALS.

The Colorectal Carcinoma (DCC) mediated attractive signalling is intrinsic to the organisation of actin cytoskeleton after activation of a complex between DCC, protein tyrosine kinase 2/FAK and src family kinases. This pathway is not reported as linked to the pathogenesis of ALS or to neurodegeneration. No protein represented in this pathway was statistically regulated in our data set.

Netrin-1 signalling includes proteins (Netrins) which are secreted and play a crucial role in neuronal migration and in axon guidance during the development of the nervous system. Among them, netrin-1 is the most studied member of the family and has been shown to play a crucial role in neuronal navigation during nervous system development mainly through its interaction with DCC and UNC5 Netrin Receptor (UNC5) receptors. Of note Fused in Sarcoma (FUS) which is known as causative protein of ALS and found to be regulated in our dataset is known to be involved in Netrin signalling with the regulation of netrin-1 splicing (195). However, none of the proteins involved in the netrin-1 signalling pathway was found to be statistically regulated in our protein list.

Oncostatin M (OSM) signalling pathway is part of the multifunctional cytokine interleukin 6 (IL6) - type cytokine family and it is mainly produced in immune cells. It is responsible of inducing transcription through different pathways, including the activation of caspase family members (CASP3, CASP7, CASP9) through the JAK2 module and has been implicated in biological mechanisms linked to apoptosis and neuronal protection (196). CASP3 regulation in our dataset was

very close to statistical significance and may be a good target for a larger scale validation experiment (unique peptide= 2, logFC= -0.708, p= 0.0544) .

CD28 dependent Vav1 pathway is involved in T-cells activation by stimulation of Vav1 through CD28, inducing multiple biochemical pathways and several cytoskeleton-dependent processes. The regulation of this pathway may bear significance in the T-cell mediated immune response known to occur in ALS (197). No protein from this pathway was found to be statistically regulated.

5.6.1 Identification of highly regulated proteins by FeaST analysis

Functional analysis of the brain/plasma NCHs proteomes provided clues on the likely biochemical changes linked to ALS and also confirmed our hypothesis that NCHs may carry important information on the pathophysiological changes occurring in this neurological disorder. For the purpose of selecting the best targets for biomarkers analysis, the list of proteins obtained using FeaST was further filtered to identify those with a high level of regulation. Statistical significance (p-value < 0.05), number of unique peptides (≥ 2) and high regulation (logFC < -0.693 OR > 0.693) were set as main criteria. This additional filtering process generated a list of 48 protein groups that were matched to the regulated pathways identified by FAT to assess their relevance in the biochemical changes identified in ALS and as biomarkers (Table 5.7).

Uniprot ID	Gene name	Protein name	Unique peptides	logFC	p-value
Q9P2W3	GNG13	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-13	2	-1.276	6.32E-03
A0A0C4DH67	IGKV1-8	Immunoglobulin kappa variable 1-8	2	-1.273	1.39E-02
O75157-2	TSC22D2	Isoform 2 of TSC22 domain family protein 2	2	-1.245	1.15E-02
Q9NRZ5	AGPAT4	1-acyl-sn-glycerol-3-phosphate acyltransferase delta	2	-1.205	6.21E-04
P35579	MYH9	Myosin-9	82	-1.000	2.70E-02
Q9Y3E2	BOLA1	BolA-like protein 1	2	-0.994	3.48E-02
Q5QJ74	TBCEL	Tubulin-specific chaperone cofactor E-like protein	2	-0.941	1.50E-02
Q53GQ0	HSD17B12	Very-long-chain 3-oxoacyl-CoA reductase	5	-0.910	7.70E-03
Q9UHI5	SLC7A8	Large neutral amino acids transporter small subunit 2	2	-0.867	2.62E-02
Q16718	NDUFA5	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 5	2	-0.862	4.52E-03
Q9BU02	THTPA	Thiamine-triphosphatase	2	-0.838	3.92E-03
O75487	GPC4	Glypican-4	2	-0.834	4.10E-02
O15084	ANKRD28	Serine/threonine-protein phosphatase 6 regulatory ankyrin repeat subunit A	3	-0.809	1.92E-02
P29144	TPP2	Tripeptidyl-peptidase 2	37	-0.789	3.62E-02
A8MWD9	SNRPGP15	Putative small nuclear ribonucleoprotein G-like protein 15	3	-0.768	1.36E-02
Q9NX63	CHCHD3	MICOS complex subunit MIC19	10	-0.763	4.14E-03

Uniprot ID	Gene name	Protein name	Unique peptides	logFC	p-value
Q53H82	LACTB2	Endoribonuclease LACTB2	2	-0.756	5.56E-03
P50238	CRIP1	Cysteine-rich protein 1	3	-0.742	2.61E-02
P00813	ADA	Adenosine deaminase	3	-0.733	1.16E-02
P51116	FXR2	Fragile X mental retardation syndrome-related protein 2	3	-0.723	1.93E-03
P40855	PEX19	Peroxisomal biogenesis factor 19	2	-0.722	9.35E-03
P38919	EIF4A3	Eukaryotic initiation factor 4A-III	6	-0.709	7.35E-03
Q9BYH1-3	SEZ6L	Isoform 2 of Seizure 6-like protein	4	-0.697	8.24E-03
Q9BV20	MRI1	Methylthioribose-1-phosphate isomerase	4	-0.692	3.74E-03
P52788	SMS	Spermine synthase	5	-0.689	2.44E-02
Q99470	SDF2	Stromal cell-derived factor 2	2	-0.682	4.26E-03
Q9NPB8	GPCPD1	Glycerophosphocholine phosphodiesterase GPCPD1	2	0.683	1.19E-02
O14514	BAI1	Brain-specific angiogenesis inhibitor 1	2	0.697	9.31E-03
Q6NXE6-2	ARMC6	Isoform 2 of Armadillo repeat-containing protein 6	2	0.700	1.87E-02
Q99729-3	HNRNPAB	Isoform 3 of Heterogeneous nuclear ribonucleoprotein A/B	3	0.706	5.29E-03
Q9HD89	RETN	Resistin	4	0.754	2.78E-02
P55083	MFAP4	Microfibril-associated glycoprotein 4	2	0.761	1.09E-02
Q06828	FMOD	Fibromodulin	3	0.770	9.28E-03

Uniprot ID	Gene name	Protein name	Unique peptides	logFC	p-value
P45984-3	MAPK9	Isoform Beta-1 of Mitogen-activated protein kinase 9	2	0.789	2.87E-02
Q9Y3C8	UFC1	Ubiquitin-fold modifier-conjugating enzyme 1	2	0.803	3.52E-02
Q9NRA0-2	SPHK2	Isoform 2 of Sphingosine kinase 2	2	0.897	3.39E-02
P04275	VWF	von Willebrand factor	127	0.910	4.69E-02
P16519-2	PCSK2	Isoform 2 of Neuroendocrine convertase 2	3	0.911	1.35E-02
P35219	CA8	Carbonic anhydrase-related protein	2	0.925	5.23E-03
Q9HAU0-2	PLEKHA5	Isoform 2 of Pleckstrin homology domain-containing family A member 5	2	0.933	1.84E-02
Q9Y328	NSG2	Neuron-specific protein family member 2	2	0.981	2.73E-03
Q00537	CDK17	Cyclin-dependent kinase 17	2	1.043	4.11E-03
O60814	HIST1H2BK	Histone H2B type 1-K	2	1.049	6.97E-03
P49406	MRPL19	39S ribosomal protein L19, mitochondrial	2	1.062	4.76E-03
Q8IX12-2	CCAR1	Isoform 2 of Cell division cycle and apoptosis regulator protein 1	2	1.113	1.77E-03
Q96F86	EDC3	Enhancer of mRNA-decapping protein 3	2	1.134	4.15E-02
O15533-2	TAPBP	Isoform 2 of Tapasin	2	1.224	1.09E-03
P27449	ATP6V0C	V-type proton ATPase 16 kDa proteolipid subunit	2	1.635	3.23E-02

Table 5.7. Protein groups highly regulated in ALS vs HC identified by FeaST analysis of TMT calibrator NCHs proteome.

The initial protein list obtained using FeaST (Paragraph 5.5) was further filtered using stringent criteria including protein groups with at least 2 unique peptides, p-Value < 0.05 and logFC < -0.693 OR > 0.693.

Uniprot ID: Uniprot database protein identifier

Gene name: the recommended gene symbol used to officially represent a gene

Protein name: protein full name recommended by Uniprot

Unique peptides: number of peptide sequences unique to a protein group

logFC: relative quantification with value expressed as $\log_2$(ALS/HC) intensities

p-value: statistical significance of the differential regulation between ALS and HC experimental groups

In the context of our study, the intersection between the most regulated proteins (obtained by Feast analysis) and the top biochemical pathways in our proteome data (identified by FAT) provides a view of the more crucial aspects of the ALS pathological cascade. In particular, four biochemical pathways seem to represent the most consistent changes at a proteomic levels: metabolism of carbohydrates (p= 0.0099), lysosome (p= 0.0015), synthesis of phosphatidic acid (p= 0.0184) and wnt signalling pathway (p= 0.0337). The main aspects of these biochemical pathways will be briefly summarized below.

5.6.1.1 Metabolism of carbohydrates

The “metabolism of carbohydrates”, a category of the Reactome database, describes the extraction of energy from sugars and its storage as glycogen. As described in Chapter 4, molecular features linked to the usage of sugars for energy production (e.g. ATP) were identified as highly represented in the HC sample. This pathway contains two highly regulated proteins: glypican-4 (GPC4) and fibromodulin (FMOD) (Table 5.7), which are also part of the glycosaminoglycans (GAGs) metabolism, a sub-pathway of “metabolism of carbohydrates” (Figure 5.9). GAGs are long, unbranched polysaccharides composed by a disaccharide unit of a hexosamine (either N-acetylgalactosamine (GalNAc) or N-acetylglucosamine (GlcNAc)) and an uronic acid (glucuronate or iduronate), which are heavily sulfated. Moreover, they are located primarily in the extracellular matrix (ECM) and on cell membranes, where the glycoproteins carrying these polysaccharides (also called proteoglycans (PGs)) are involved in structural and morphological changes but also in cell signaling (198). PGs are produced in the Golgi and degraded in the lysosome (Figure 5.9). In particular, FMOD and GPC4 are part of protein complexes involved in keratan sulphate (KS) and heparan sulphate (HS) PGs synthesis and degradation.

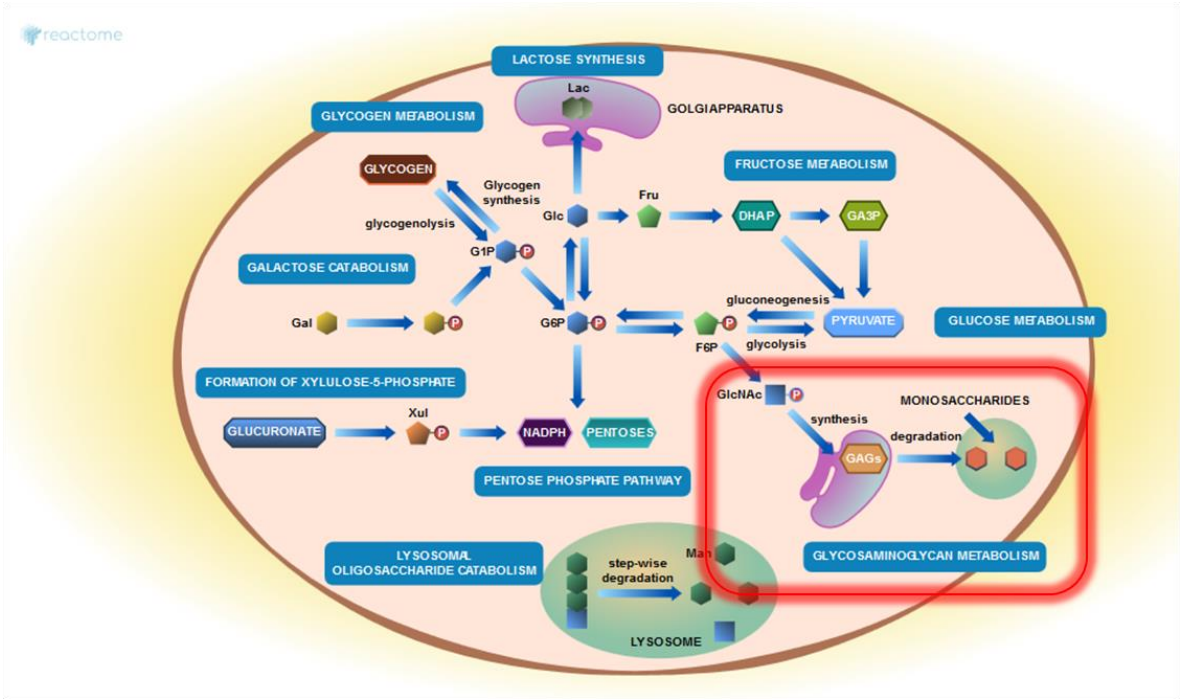


Figure 5.9. Metabolim of carbohydrates: Reactome database.

The metabolism of carbohydrates is a complex interplay of pathways involved in energy production from sugars and its storage as ATP and glycogen. The red box highlight glycosamminoglycan metabolism. Image source https://reactome.org/PathwayBrowser/.

Keratin sulphate (KS) is linked to several proteins that are expressed in the extracellular matrix (ECM) of connective and epithelial tissues, but also in the nervous system (199). This protein has been implicated in the pathogenesis of ALS, with the report of changes of KS sulfonation levels of spinal cord microglia in SOD1G93A mice and ALS patients in the early and late phase of the disease (200,201).

Proteoglycans carrying heparan sulphate (HS, HSPGs) are ubiquitously expressed and due to their location can function as receptors, which are possibly involved in aggregation processes (202,203). In particular, it has been shown that HSPGs are mediators of tau fibril binding and uptake in this protein' seeding of aggregation in mouse and human cell lines (204). HS branches were shown to be responsible of fibril formation, acting as a scaffold that stabilises the oligomeric proto-filament structure. This fibril formation process appears to be relevant also for prion protein, islet amyloid polypeptide, a-synuclein, amyloid-ß and tau (205). It is therefore possible that that GAGs, in particular HSPG, could play a central role in protein aggregation biology that is a key pathological feature in ALS.

The review of our proteomic data on the GPC4 protein complex showed that biglycan (BGN; unique peptides= 8; logFC= -0.369; p= 0.024) was significantly regulated, adding substance to the potential role for GAGs metabolism involvement in the biochemical changes observed in ALS. Biglycan (BGN) is a protein involved in HSPGs synthesis and in chondroitin sulphate (CS) PGs synthesis and degradation. Chondroitin sulphate proteoglycans (CSPGs) are the major component of ECM in mammal's central nervous system. Critically for ALS pathogenesis and neurodegeneration in general, an abnormal expression of CS proteoglycans has been shown in astrocytes (206) and spinal motor neurons (207) of SOD1 transgenic rats, while CSPGs have been found inside of inclusion bodies in brain tissues from patients affected by several neurodegenerative diseases (208).

5.6.1.2 Lysosome

The KEGG database "lysosome" pathway encompasses the biochemical changes taking place during lysosomes formation. Lysosomes are membrane-delimited organelles working as the main digestive compartment within the cells, thanks to the presence of hydrolytic enzymes that can break down different molecules. After synthesis in the endoplasmatic reticulum (ER), lysosomal enzymes are decorated with mannose-6-phosphate residues, which are recognized by mannose-6-phosphate receptors (M6PR and MPRI) in the trans-Golgi network, then packaged into clathrin-coated vesicles and transported to late endosomes to fuse into the lysosome (Figure 5.10). Molecules are acquired for digestion in the lysosome by endocytosis, phagocytosis and autophagy (Figure 5.10).

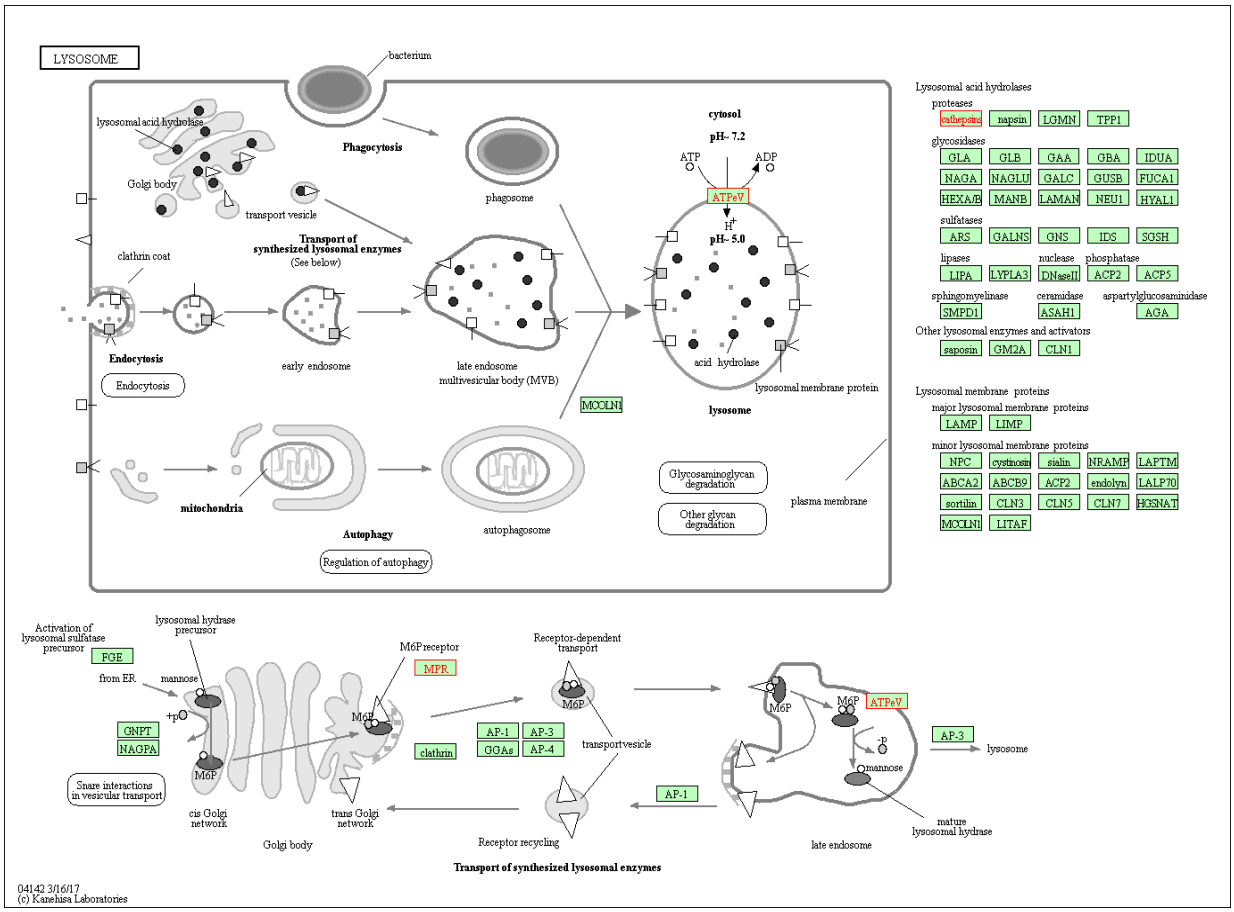


Figure 5.10. Lysosome pathway: KEGG database.

The figure depicts the entire KEGG lysosome pathway with squared rectangles representing gene products (e.g. proteins) and rounded rectangles representing maps (e.g. other pathways). The objects filled in green indicated the genes/proteins involved in the pathway, while the boxes highlighted with a red outline and font indicated the genes/proteins significantly regulated in our data set.

In our proteome, the lysosome pathway emerges as relevant at different levels: 1) the regulation of key proteins like the V-type proton ATPase 16 kDa proteolipid subunit (ATP6V0C) (Table 5.7) , of cathepsin L1 (CTSL; unique peptides= 2, logFC= -0.529, p= 0.0115) and Cation-dependent mannose-6-phosphate receptor (M6PR; unique peptides= 2, logFC= 0.522, p= 0.0384) and 2) the regulation of the GO cellular component (CC) lysosome function-related clathrin-coated vesicle membrane (p= 0.00751) and biological process (BP) adaptor protein (AP)-2 adaptor complex (p= 0.0115) emerging through FAT analysis.

Lysosomal dysfunction is well described in ALS especially in relation with autophagy (autolysosome). It is not clear whether the lysosomal proteolytic cascade is synergic to or in opposition to activity of the proteasome, raising

questions regarding how protein degradation in ALS become affected in the pathological process (150,151). In sporadic ALS, it has been shown that molecular pathways driving autophagy are activated specifically in motor neurons (209). Another important evidence of the involvement of autophagy in ALS comes from studies on the main ALS causative gene, C9orf72, which is supposed to play a central role in vesicle trafficking and autophagy initiation (210,211).

ATP6V0C, a highly regulated protein identified in this pathway, is a proton-conducting pore subunit of the membrane integral V0 complex of vacuolar ATPase, which is responsible for lysosome acidification. It has also a central role in cell homeostasis as part of gap-junction and neurotransmitter release when it is located in the plasma membrane (212,213). Its inhibition leads to apoptosis (214).

CTSL is part of a class of proteins called cathepsins, a group of lysosomal proteases that are activated by the low pH and have a key role in cellular protein turnover. Cathepsins include different type of proteases (serine, aspartic and cysteine proteases) and CTSL is part of the cysteine protease. The hydrolytic activity of these proteins in is reduced in ALS and in other neurodegenerative diseases (215–217). Plasma levels of cathepsins in patients affected by Alzheimer's disease were found to be higher than in HC but not specifically CTSL (218). Interestingly, through the activation of CASP3, CTSL has a role in the cross-talk between autophagy and apoptosis in neuronal cell line exposed to dopamine (DA) (219). This connection could be of relevance in our study, since DA release cycle is one of the key pathways in our dataset (Table 5.6). Moreover, it has been shown that in a catechol-O-methyltransferase (COMT) drosophila model, a defect in CTSL trafficking to the lysosome leads to inability to sustain autophagy under stress (220) , adding evidence that autophagy impairment and CTSL trafficking are centrally placed in the pathogenesis of neurodegenerative disorders.

The other differentially regulated protein found in this pathway was M6PR, which is involved in the transport of phosphorylated lysosomal enzymes from the Golgi complex and from the cell surface to lysosomes. Lysosomal enzymes bearing phosphomannosyl residues bind specifically to mannose-6-phosphate receptors

in the Golgi apparatus and the resulting receptor-ligand complex is transported to an acidic prelysosomal compartment where the low pH mediates the dissociation of the complex. This process was found to be impaired in a model of Parkinson's disease suggesting that the altered lysosome activity may be a shared pathological feature across neurodegenerative diseases (221).

5.6.1.3 Synthesis of phosphatidic acid

The Reactome database "synthesis of phosphatidic acid (PA)" is about the de novo synthesis of PA (Figure 5.11). The pathway starts from the esterification of glycerol-3-phosphate (G3P) to lysophosphatidic acid (LPA) mediated by glycerol 3-phosphate acyltransferase (GPAT), and then LPA is converted to PA by a LPA acyltransferase (AGPAT, also known as LPAAT) (Figure 5.12). PA can be also formed by phospholipases through hydrolysis of glycerophospholipids, such as phosphatidylcholine (PC) and phosphatidylinositol (PI) (Figure 5.11). Because PA is a key intermediate metabolite in the synthesis of membrane glycerophospholipids, it is also involved in acyl chain remodelling in plasma membrane via cleavage by phospholipases followed by re-acylation by acyltransferases (222–226). Acyl chains are responsible of membrane curvature playing a significant role in membrane vesicle fission and fusion events (227,228). The relevance of this pathway in our proteomic analysis relates also to the presence of regulated genes like AGPAT (Table 5.7) and lysophosphatidylcholine acyltransferase 1 (LPCAT1; unique peptides 1, logFC -0.622, p= 0.0156).

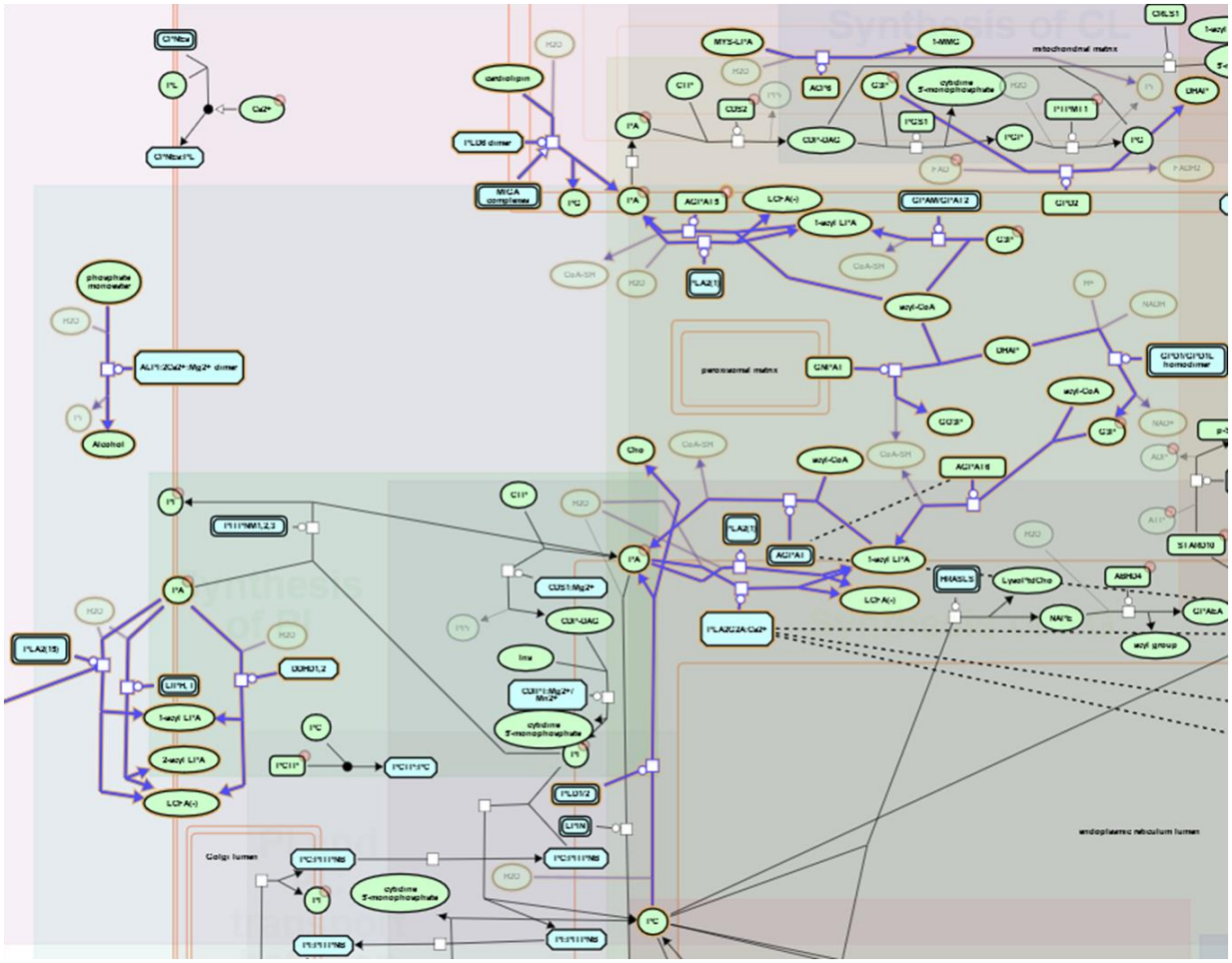


Figure 5.11. Synthesis of phosphatidic acid (PA): Reactome database.

The figures shows the biochemical reactions in the synthesis of phosphatidic (PA), which is an intermediate of the glycerophospholipids metabolism. The boxes with brighter colours and the blue arrows show the reactions and the molecules taking part in PA synthesis. Green rounded circles indicate small molecules (e.g. PA), green rounded boxes indicate proteins, while cyan boxes indicate complexes.

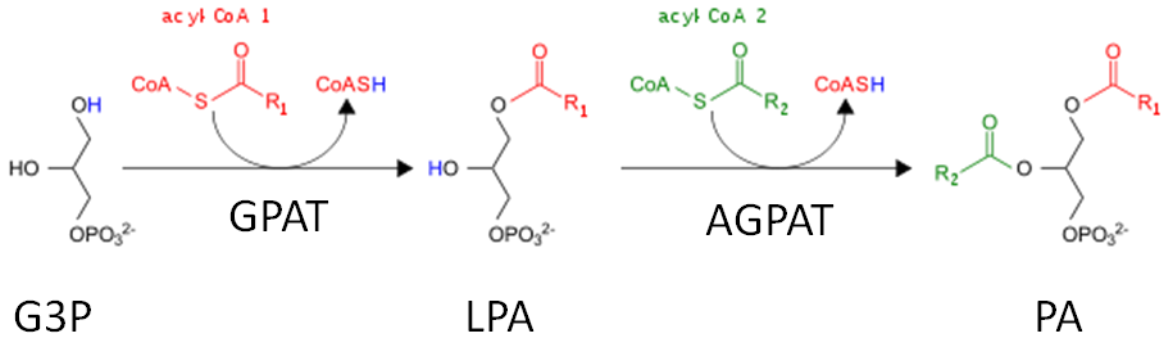


Figure 5.12. Phosphatidic acid synthesis from glycerol-3-phosphate.

Glycerol-3-phosphate (G3P) is converted into lysophosphatidic acid (LPA) through esterification by the glycerol 3-phosphate acyltransferase (GPAT), followed by another esterification by LPA acyltransferase (AGPAT) to produce phosphatidic acid (PA). AGPAT consists of a complex of ten

enzymes. The two acyl chains (R_1 and R_2) are transferred to G3P and LPA from two molecules of acyl-coenzyme A (acyl-CoA1 and 1). Depending on the structure and length of R_1 and R_2 it can be later converted to phosphatidylcholine (PC) and phosphatidylinositol (PI).

PA has not been directly linked to ALS; however it is synergic with LPA function which is involved in cell signalling pathways through G-protein-coupled receptors (GPCRs) and Ca^{2+} level intracellularly which can ultimately be detrimental to cell function (229). A further way in which LPA can modulate cell behaviour is by altering membrane dynamics and structure. It can stabilise the plasma membrane bilayer or induce monolayer curvature by endophilin or 50 kDa brefeldin A ribosylated substrate (BARS-50) activity, affecting endocytic and Golgi membrane fission (230). Endophilin-B2, identified as secondary isoform, was also found significantly regulated (SH3GLB2; unique peptides= 11, logFC= 0.340, p= 0.0328).

PA is also functionally linked to all types of phospholipids, biomolecules that have been described in ALS pathology. Phosphatidylcholine (PC) was found as highly discriminant of ALS pathology in a lipidomic study comparing CSF from ALS patients to a group of controls (231), while phosphatidylethanolamine (PE) has been evaluated as a cofactor in prion propagation in vitro (232). Glycerophospholipids are also involved in cell signaling, as their presence at the membrane surface in neurons can be a signal for phagocytosis and a marker for neuroinflammation, as seen in Alzheimer's and Parkinson's diseases (233,234). This suggests that an imbalance in PA (or of glycerophospholipids linked to PA) biosynthesis can have a role in neurodegenerative processes. To date, there has not been any published comprehensive evaluation of phospholipid expression in plasma from ALS patients.

5.6.1.4 Wnt signalling pathway

Wnt (wingless/integrated, named after studies conducted in Drosophila) proteins are secreted morphogens that are required for basic developmental processes, such as cell-fate specification, progenitor-cell proliferation and the control of asymmetric cell division, in many different species and organs (235–239). There are at least three different wnt pathways: the canonical pathway, the planar cell polarity (PCP) pathway and the wnt/Ca2+ pathway (Figure 5.13). The highly regulated protein found in this pathway is GPC4 (Table 5.7), which was also

involved in GAGs metabolism. Serine/threonine-protein phosphatase 2B catalytic subunit alpha (PPP3CA; unique peptides=1, logFC=0.983, p= 0.0147) was also found as regulated in this pathway. GPC4 takes part in PCP signalling, leading to the activation of the small GTPases RHOA (RAS homologue gene-family member A) and RAC1 (Ras-related C3 botulinum toxin substrate 1), which activate the stress kinase JNK (Jun N-terminal kinase) and ROCK (RHO-associated coiled-coil-containing protein kinase 1) and leads to remodelling of the cytoskeleton and changes in cell adhesion and motility (Figure 5.13). PPP3CA is part of the wnt-Ca^{2+} signalling mediated through G proteins and phospholipases and leads to transient increases in cytoplasmic free calcium that subsequently activate the protein kinase C (PKC) and calcium calmodulin mediated kinase II (CAMKII) and the phosphatase calcineurin (Figure 5.13).

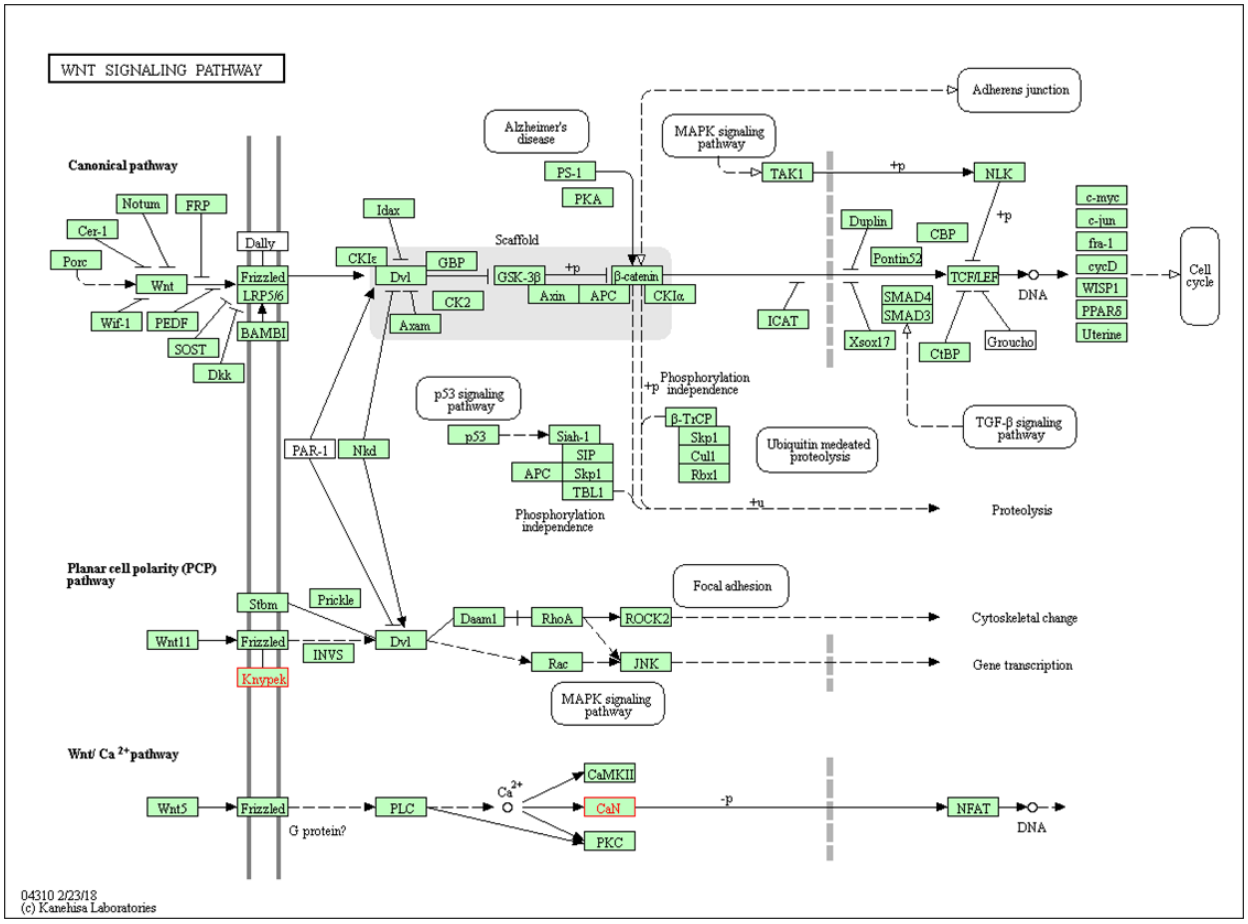


Figure 5.13. Wnt signalling pathway: KEGG database.

The figure reports the KEGG pathway with squared rectangles representing gene products (e.g. proteins) and rounded rectangles representing maps (e.g. other pathways). The objects filled in green indicated the genes/proteins involved in the pathway in Homo sapiens and the boxes in red

outline and font indicate the genes/proteins significantly regulated in our data set. Knypek is GPC4, while CaN is the protein phosphatase 3 (PP3) complex including also PP3CA.

Published data corroborates wnt signalling pathway involvement in ALS as a downstream effect of the neurodegeneration process and inflammation seen in SOD1G93A transgenic mice (240–242). However, wnt, and other proteins involved in this pathway were not significantly regulated in our data set.

5.6.2 Other pathways of relevance in ALS

Looking at the regulated pathway obtained from FAT analysis and matching ALS literature, it is worth mentioning: synaptic vesicle pathway (p= 0.00881), neurotransmitter release cycle (p= 0.0179), translocation of GLUT4 to the plasma membrane (p= 0.0191), SUMOylation of chromatin organisation proteins (p= 0.0234) and SUMOylation of RNA binding proteins (p= 0.0247).

Synaptic vesicle pathway, from the Wikipathways database, embodies the transmission of neurotransmitters from the presynaptic nerve to the synaptic cleft. Two proteins from this pathway were significantly regulated in our data set, protein/nucleic acid deglycase DJ-1 (PARK7; unique peptides= 19, logFC= 0.445, p= 0.00666) and sodium/potassium-transporting ATPase subunit α2 (ATP1A2; unique peptides= 35, logFC= 0.271, p= 0.0475). In particular, PARK7 was found to affect disease progression rate in PARK7 knock-out $SOD1^{G93A}$ mouse model of ALS accelerating the disease course (243), while PARK7 mRNA was found to be down-regulated in muscle tissue from $SOD1^{G93A}$ mice models of ALS (244). PARK7 may be involved in protein aggregation for its deglycase activity. Glycation and advanced glycation end-products (AGEs) have been found in protein aggregates as PTM in the context of studies on neurodegenerative diseases (245–248).

Neurotransmitter release cycle (Reactome database) describes the multi-step process of synaptic vesicle progress into the synaptic cleft, upon depolarisation of the pre-synaptic membrane. Two proteins taking part in this pathway were significantly regulated in our data set: amine oxidase A (MAOA; unique peptides= 10, logFC= 0.319, p= 0.0491) and LIN7B. However, none of these proteins have been previously linked with ALS pathophysiology.

Translocation of GLUT4 to the plasma membrane (Reactome database) describes the translocation of the glucose transporter GLUT4 to the plasma membrane, by intracellular vesicle, for glucose uptake from the bloodstream. This pathway has been recently described in plasma samples from ALS patients and $SOD1^{G93A}$ mice as the possible connection between the lactate, produced as the by-product of the cell metabolism, and chronic inflammation casuing neuromuscular junction disruption and motor neurons death (249). One protein has been found statistically regulated, ras-related protein rab-10 (RAB10; unique peptides= 7, logFC= -0.299, p= 0.0465) that has been associated with Parkinson's disease (250).

SUMOylation is a type of post translational modification (PTM) involved in several cellular processes, where a small ubiquitin-like modifier protein (SUMO) is attached or detached from a protein. The pathways, SUMOylation of chromatin organisation proteins and SUMOylation of RNA binding proteins (both from the Reactome database) cover the steps in which this PTM modifies chromatin organisation and alters RNA binding proteins interactions with nucleic acids. In these two pathways, only the small ubiquitin-related modifier 2 (SUMO2; unique peptides= 1, logFC= 1.127, p= 0.0138) was statistically regulated in the data set. SUMOylation has been described in ALS in the context of studies which have looked at protein aggregation (251,252).

5.7 Validation of selected candidate biomarkers in plasma NCHs

Taking together mass spectrometry data, the levels of regulation, the statistical significance and the role within pathway involved in ALS pathology, six proteins were selected for further validation by western blot (Table 5.8).

Gene name	Protein name	LogFC	Pathway
FMOD	Fibromodulin	0.77	Glycosaminoglycans (GAGs) metabolism (Keratan sulfate)
GPC4	Glypican-4	-0.834	GAGs metabolism (Heparan sulfate)
BGN	Biglycan	-0.369	GAGs metabolism (Chondroitin/Dermatan sulfate)

Gene name	Protein name	LogFC	Pathway
M6PR	Cation-dependent mannose-6-phosphate receptor	0.522	Lysosome
SH3GLB2	Endophilin-B2	0.34	Synthesis of PA
PARK7	Protein DJ-1 (or Protein/nucleic acid deglycase DJ-1)	0.445	Synaptic Vesicle Pathway

Table 5.8. Biomarker candidates selected for further validation by Western Blot.

Protein candidates extracted from regulated pathways and previously linked to ALS pathogenesis are selected for further validation by immunodetection.

Gene name: the recommended gene symbol used to officially represent a gene

Protein name: protein full name recommended by Uniprot

logFC: relative quantification with value expressed as $\log_2$(ALS/HC) intensities

Pathway: name of the pathway in which the protein was identified

The validation was carried out on plasma NCHs extracted from ALS patients and HC (Table 5.9; Figure 5.14-Figure 5.19).

Sample ID	Gender	Ethnicity	El-Escorial	Clinical Onset	Age at visit (year)	ALSFRS-R	Rate of progression at visit
ALS1*	F	Caucasian	Definite ALS	LL	61,11 - 61,99	19 – 36	0,649 - 0,997
ALS2*	F	Caucasian	Definite ALS	UL	68,45 - 68,8	35	0,821 - 1,111
ALS3*	M	Caucasian	Definite ALS	LL	59,62 - 61,18	35 – 38	0,436 – 0,853
ALS4*	M	Caucasian	Definite ALS	UL	64,40 - 65,82	8 – 12	1,423 – 2,369
HC3	M	Caucasian	NA	NA	60.64	NA	NA
HC6	F	Caucasian	NA	NA	63.53	NA	NA
HC7	F	Caucasian	NA	NA	60.3	NA	NA
HC8	F	Caucasian	NA	NA	66.4	NA	NA
HC9	F	Caucasian	NA	NA	63.1	NA	NA
HC10	M	NA	NA	NA	56.6	NA	NA

Table 5.9. Demographic and clinical features of ALS patients and HC selected for the validation of six relevant protein biomarkers in plasma NCHs.
In light orange background the Amyotrophic Lateral Sclerosis (ALS) patients and in light blue background the healthy control (HC). If the enriched fractions were obtained using more than one time point plasma sample for a single patient, the range values for age at visit, ALSFRS-R and progression rate were reported. HC3 and HC6 are the same samples listed in Table 5.1; *: samples obtained from the same ALS patients listed in Table 5.1, but at different time points.
Sample ID: ALS= ALS patients; HC= Healthy Control

Gender: M= Male; F= Female

El-Escorial: diagnostic classification of ALS diagnosis (88)

Clinical Onset (site of initial clinical signs): LL= Lower Limbs; UL= Upper Limbs

Age at sampling: age of the patient at blood sampling

ALSFRS-R (ALS Functional Rating Scale revised): level of neurological impairment across different clinical domains (1-48, with higher neurological impairment with lower values)

Rate of progression at sampling: calculated as 48 - ALSFRS-R score at sampling time/disease duration from onset of symptoms to sampling time

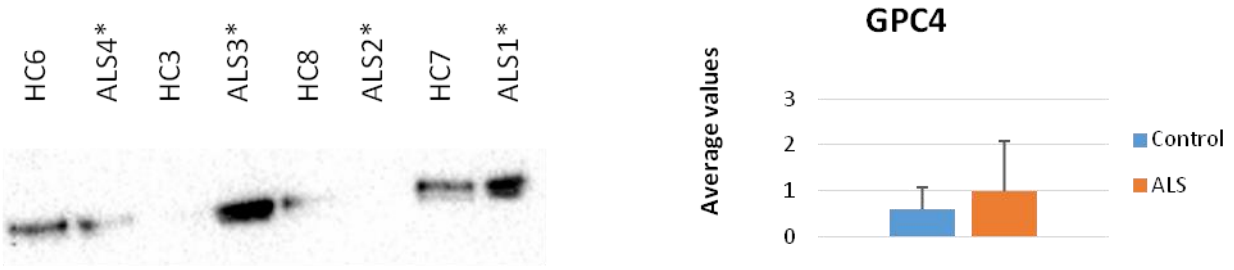


Figure 5.14. Western blot analysis of Glypican-4 (GPC4) in plasma NCHs from ALS patients and healthy controls.

The image was acquired with a Chemi-Doc Camera (Bio-Rad), while band densities analysed by Image Lab software (Bio-Rad). Samples were normalised to HC6 density and the average values, with relative standard deviation, for the ALS (n=4) and Control (n=4) groups were plotted onto the chart. The immunodetection confirmed the presence of the protein target at the expected molecular weight in plasma NCHs from ALS and HC samples, but contrary to what shown by the TMTcalibrator™ experiment, with a higher content in the ALS group (logFC= -0.834).

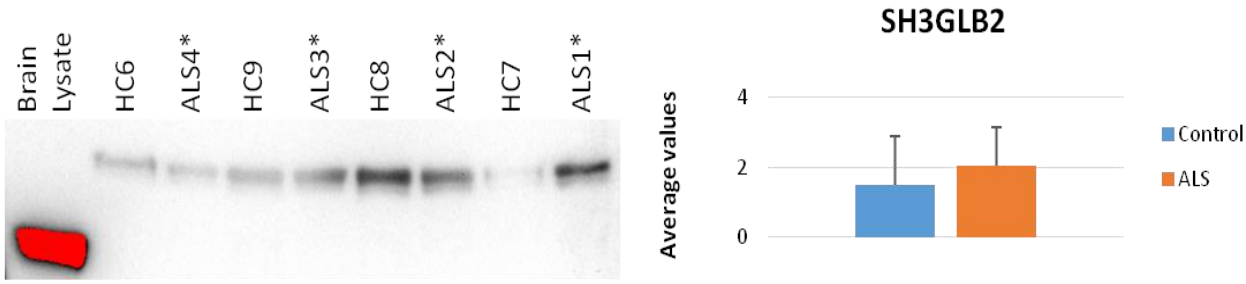


Figure 5.15. Western blot analysis of Endophilin-B2 (SH3GLB2) in plasma NCHs from ALS patients and healthy controls.

The image was acquired with a Chemi-Doc Camera (Bio-Rad), while band densities analysed by Image Lab software (Bio-Rad). Samples were normalised to HC6 density and the average values with relative standard deviation for the ALS (n=4) and Control (n=4) groups were plotted onto the chart. A brain lysate sample was also included (red band, indicating signal saturation) which showed an endophilin-B2 band at a lower molecular weight than the bands detected in plasma NCHs. In line with the TMTcalibrator™ experiment data, the immunodetection confirmed the presence of the protein target in plasma NCHs from ALS and HC samples with a higher level of expression in the ALS group (logFC= 0.34).

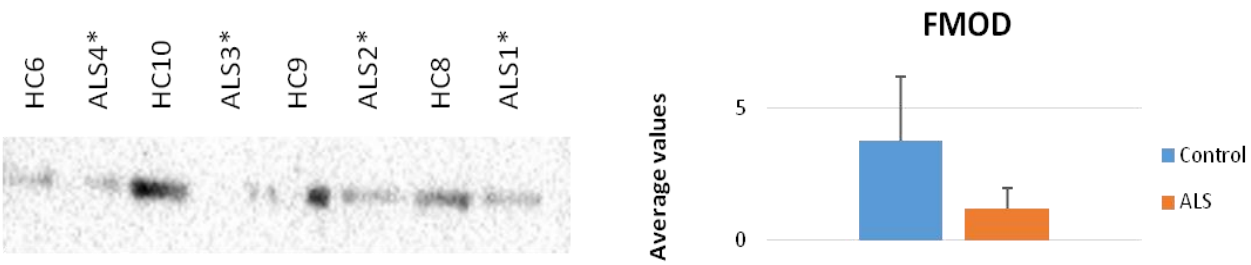


Figure 5.16. Western blot analysis of Fibromodulin (FMOD) in plasma NCHs from ALS patients and healthy controls

The image was acquired with a Chemi-Doc Camera (Bio-Rad), while band densities analysed by Image Lab software (Bio-Rad). Samples were normalised to HC6 density and the average values with relative standard deviation for the ALS (n=4) and Control (n=4) groups were plotted onto the chart. The immunodetection confirmed the presence of the protein target at the expected molecular weight in plasma NCHs from ALS and HC samples, but contrary to what shown by the TMTcalibrator™ experiment, with a higher expression in the Control group (logFC= 0.77).

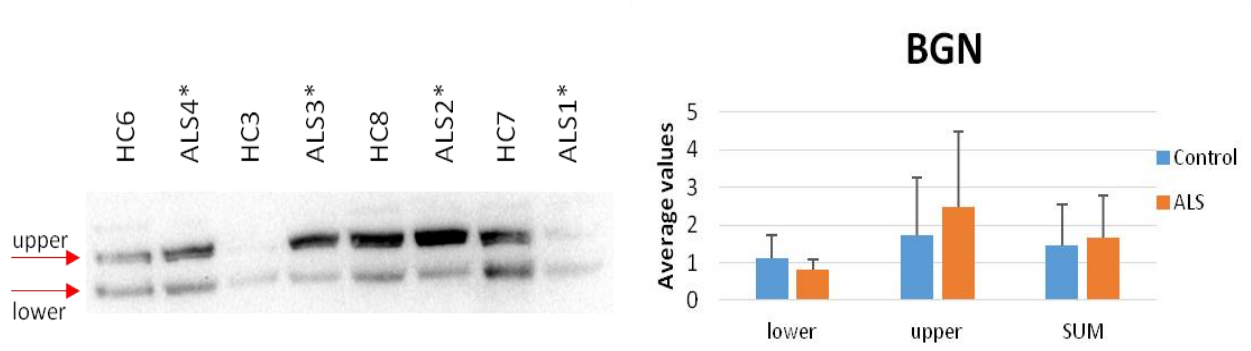


Figure 5.17. Western blot analysis of Biglycan (BGN) in plasma NCHs from ALS patients and healthy controls

The image was acquired with a Chemi-Doc Camera (Bio-Rad), while band densities analysed by Image Lab software (Bio-Rad). Samples were normalised to HC6 density and the average values with relative standard deviation for the ALS (n=4) and Control (n=4) groups were plotted onto the chart. The immunodetection revealed two bands, one at 49KDa (lower) and the other one at 62KDa (upper) in line with the manufacturer (Cambridge Bioscience) data on U-138 MG cell line lysate. The relative amount of the two bands showed a different trend for the lower and the upper bands across patients and controls as shown on the graph on the right-hand side. Contrary to what shown by the TMTcalibrator™ experiment (logFC= -0.369), the sum of the densitometric values for each individuals showed a slightly higher total BGN signal in the ALS group (1.47 for Control and 1.67 for ALS).

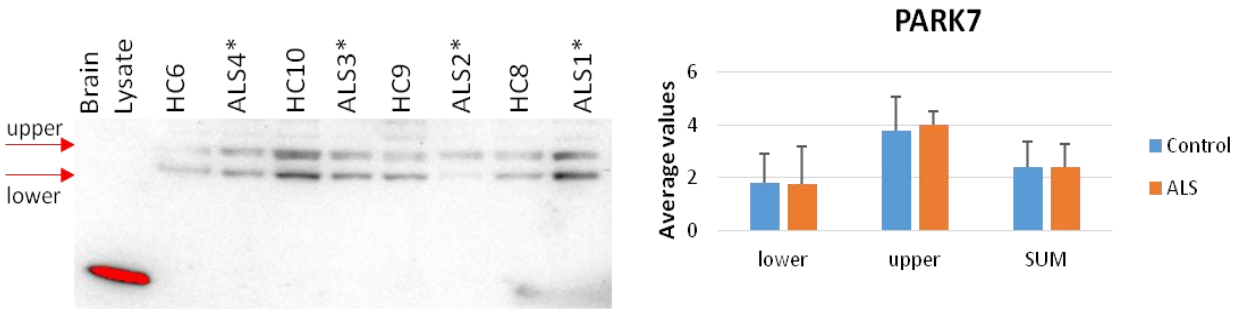


Figure 5.18. Western blot analysis of protein/nucleic acid deglycase DJ-1 (PARK7) in plasma NCHs from ALS patients and healthy controls.

The image was acquired with a Chemi-Doc Camera (Bio-Rad), while band densities analysed by Image Lab software (Bio-Rad). Samples were normalised to HC6 density and the average values with relative standard deviation for the ALS (n=4) and Control (n=4) groups were plotted onto the chart. The immunodetection revealed two bands for plasma aggregates samples at 49 and 62KDa (defined in the graph as lower and upper respectively), while according to the Manufacturer (Cambridge Bioscience) one band only at 25KDa is supposed to be detected. A brain lysate sample included in the western blot (red band, indicating saturation) showed only one band at 28KDa, suggesting that the western blot probing conditions were specific for the PARK7 protein and that the higher than expected molecular weight band identified in the plsma NCHs could be related to the formation of aggregated forms. Contrary to what shown by the TMTcalibrator™ experiment (logFC= 0.445), the densitometric signal was almost the identical for the lower and upper band and also for the sum of the two.

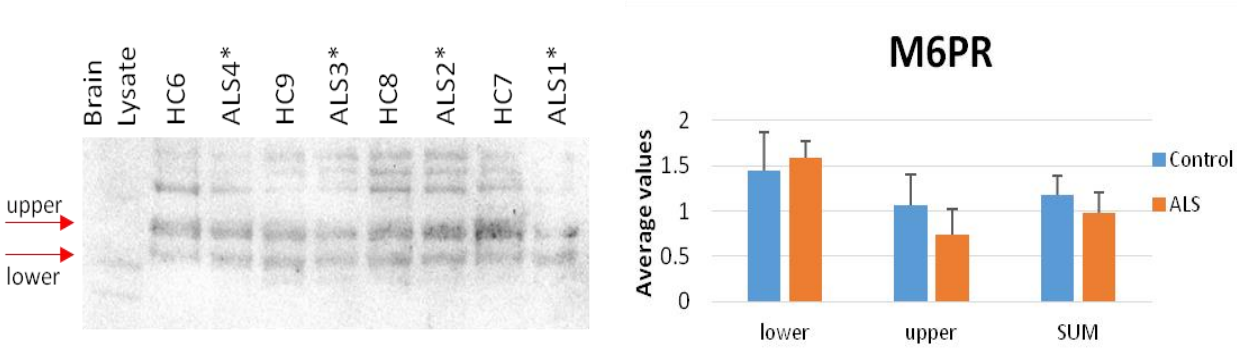


Figure 5.19. Western blot analysis of Cation-dependent mannose-6-phosphate receptor (M6PR) in plasma NCHs from ALS patients and healthy controls.

The image was acquired with a Chemi-Doc Camera (Bio-Rad), while band densities analysed by Image Lab software (Bio-Rad). Samples were normalised to HC6 density and the average values with relative standard deviation for the ALS (n=4) and Control (n=4) groups were plotted onto the chart. The immunodetection revealed several bands for plasma aggregates samples with the two most intense at 49 and 62KDa (defined in the graph lower and upper respectively), while Manifacturer (Insight Biotechnology) reports a detection range of ≈35 and ≈60KDa. A brain lysate sample included in this experiment showed two bands at 38 and 49KDa, suggesting that the probing conditions could be specific for the M6PR protein. To quantify the signal, taking into

account only the bands at 49 and 62KDa, the relative intensityof the two bands showed a different trend for the lower and the upper bands as shown in the figure, while the sum of the two bands intensities was slightly higher in the Control group, contrary to what shown by the TMTcalibrator™ experiment data (logFC= 0.522).

From the six candidates tested on plasma NCHs, endophilin-B2 (SH3GLB2) was the only protein that showed the same trend of protein regulation comparing ALS vs HC identified in the TMTcalibrator™ experiment, with a similar level of regulation (logFC= 0.34 in TMTcalibrator™ and $\log_2$(ALS/HC)= 0.437 in WB analysis), while the other candidates showed an opposite trend of regulation (GPC4, FMOD, BGN and M6PR) or no regulation at all (PARK7). However, it is worth reporting that GPC4 (Figure 5.14) and FMOD (Figure 5.16), still showed a significant level of differential expression between ALS and HC (in particular FMOD that was three-fold more expressed in HC compared to the ALS, while GPC4 was almost two-fold higher in ALS than HC).

Interestingly, three protein candidates (endophilin-B2, PARK7 and M6PR) (Figure 5.15, Figure 5.18 and Figure 5.19) were detected at a higher MW in plasma NCHs than in brain lysate. This difference could be related to different PTM patterns encountered in tissues as opposed to fluids and to the state of aggregation. In fact, Uniprot database confirmed the presence of several PTMs for these three proteins. PARK7 can be subjected to lipidation with palmitoyl in three residues, while M6PR can be glycosilated in five different residues and endophilin-B2 can be phosphorylated in two residues. The different tissues and sample type (lysate and aggregates) could justify different level of modification and so the final molecular weight at which the proteins were detected. The same could be applied to BGN where the two bands obtained in the WB (Figure 5.17) could represent different glycosylation levels in six different residues.

The potentially high level of PTMs that is possible to encounter in protein aggregates, especially those coming from blood that can form AGEs, in addition to the changes occurring through regulated enzymatic processes like the effect of proteases could explain the different level of regulation between TMTcalibrator™ and WB data. Unfortunately, this study was conducted without including any biological PTM (e.g. phosphorylation) during PSM search of the output MS data after spectra acquisition, making impossible to evaluate the NCH-proteome in a more accurate manner. On the other hand, including these PTMs

without specific sample preparation for peptide enrichment would have given an inaccurate quantification and regulation. Enrichment of peptides carrying a specific PTM requires a higher amount of starting material, which is a limiting factor in any NCHs enrichment process (more than 2 ml of starting plasma material per sample only for mass spectrometry analysis would be needed). The high volume of specimen required is probably the main limit associated to the study of protein aggregates.

Another important factor that may affect the level of protein regulation in ALS and in HC, which is investigated using two different methodologies (TMTcalibrator™ and WB), is probably intrinsic to the TMTcalibrator™ bioinformatic pipeline. In fact for both experimental groups (ALS and HC), the workflow for quantification analysis considers only the peptides detected in the calibrant channels (in our case brain) that are also identified in the analytical channels (in our case plasma NCHs). This means that a number of peptides derived from our selected candidates, which may be constitutively expressed only in plasma and not in brain, may escape quantification and downstream functional analysis in the TMT bioinformatic process of quantitation. On the contrary, immunodetection in WB may detect the whole spectrum of peptides (both from plasma and from brain) giving a different final result in terms of overall quantification compared to TMT calibrator. In addition, proteins detected in brain lysate (Figure 5.15, Figure 5.18 and Figure 5.19) may have a different MW suggesting different PTM patterns, adding an additional bias and confounding factor in the levels of protein regulation obtained in the proteomic experiment.

5.8 Validation of selected candidate biomarkers in neat and depleted plasma

In order to evaluate the potential of the selected protein as biomarkers, the same validation in ALS vs HC was extended to neat plasma and to albumin and IgGs depleted plasma. We have tested the hypothesis that direct plasma analysis without having to rely on the process of NCHs extraction would simplify the analysis, improving rapidity and high throughput and eventual translation into diagnostic tools in clinical application. This second round of validation was carried on a different set of ALS patients and HCs (Table 5.10).

The presence of the six biomarker candidates (FMOD, GPC4, BGN, PARK7, SH3GLB2 and M6PR) was evaluated by western blot loading 80 µg of proteins for each depleted plasma sample and 100 µg for each neat plasma sample (Figure 5.20-Figure 5.25). In order to evaluate the relative quantification for protein expression, we considered to include an internal control in western blot analysis. However, it has been shown that ALS patients have an imbalanced production for all the proteins suggested as internal controls in blood plasma, such as albumin, IgGs and ferritin. For this reason, no internal loading control was used in these experiments.

Sample ID	Gender	Ethnicity	El-Escorial	Clinical onset	Age at visit (years)	ALSFRS-R	Rate of progression
A1	M	Caucasian	suspected ALS	L(LL)	65.2	39	0.335
A2	M	Asian	definite ALS	R and LL	61.9	36	2.666
A3	F	Caucasian	probable ALS	L(LL)	69.5	37	0.257
A4	F	Caucasian	definite ALS	B	68.4	38	0.89
A5	M	Caucasian	possible ALS	B	67	43	0.365
A6	M	Caucasian	probable ALS	B	82.5	27	2.096
C1	M	Caucasian	NA	NA	63.9	NA	NA
C2	F	Caucasian	NA	NA	65.9	NA	NA
C3	M	Caucasian	NA	NA	68.3	NA	NA
C4	M	Caucasian	NA	NA	73.1	NA	NA
C5	M	Caucasian	NA	NA	60.3	NA	NA

Table 5.10. Demographic and clinical features of ALS patients and healthy controls selected for the validation of six relevant protein biomarkers in neat and depleted plasma.

In light orange colour the Amyotrophic Lateral Sclerosis (ALS) patients and in light blue background the healthy control (HC).

Sample ID: ALS= ALS patients; HC= Healthy Control

Gender: M= Male; F= Female

El-Escorial: diagnostic classification of ALS diagnosis (88)

Clinical Onset (site of initial clinical signs): L= Limbic; LL= Lower Limbs; B= Bulbar; R= Respiratory

Age at sampling: age of the patient at blood sampling

ALSFRS-R (ALS Functional Rating Scale revised): level of neurological impairment across different clinical domains (1-48, with higher neurological impairment with lower values)

Rate of progression at sampling: calculated as 48 - ALSFRS-R score at sampling time/disease duration from onset of symptoms to sampling time

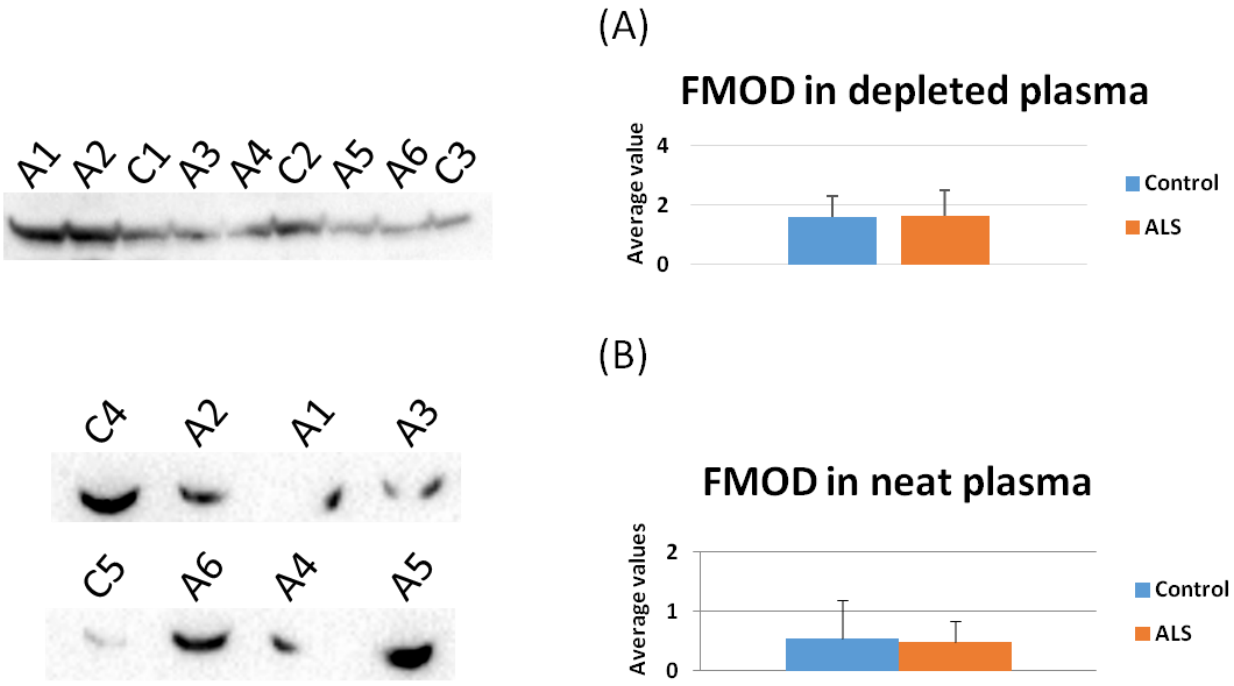


Figure 5.20. Immunodetection of FMOD in depleted and neat plasma.

(A) and (B) show immunodetection and relative quantification, of FMOD in samples listed in Table 5.10. Quantification was performed normalising on C3 for (A) and C4 in (B).

FMOD was detected in both plasma sample types depleted (Figure 5.20A) and neat (Figure 5.20B) plasma. C5, A1 and A4 did not show a complete band profile. Quantification of the blots showed that FMOD is present in plasma and depleted plasma at the same concentration for ALS and Control groups. This result is in contrast to the analyses in NCHs, with the TMTcalibrator™ showing higher FMOD expression in the ALS samples (Table 5.8) and the immunodetection in which FMOD was higher in the Control group (Figure 5.16).

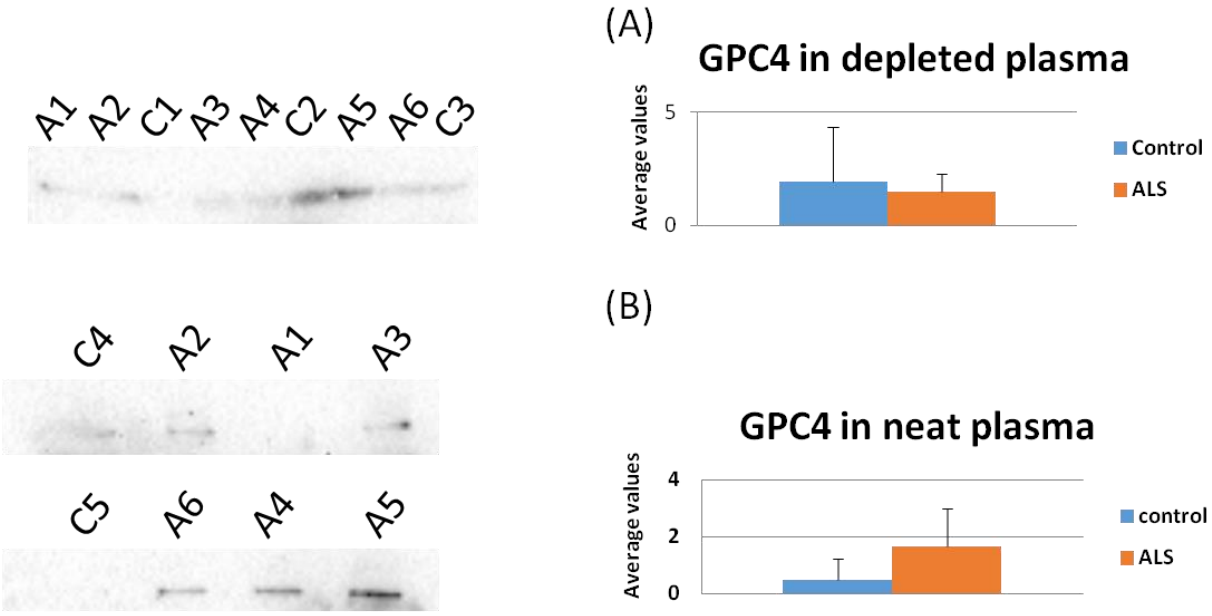


Figure 5.21. Immunodetection of GPC4 in depleted and neat plasma.

(A) and (B) show immunodetection and relative quantification of GPC4 in samples listed in Table 5.10. Quantification was performed normalising on C3 for (A) and C4 in (B).

Immunodetection of GPC4 was difficult, in fact the bands detected in the depleted plasma (Figure 5.21A) were very faint, while in neat plasma (Figure 5.21B) two samples, C5 and A1 did not show any band. Quantification of the target was not consistent in the two sample types (Figure 5.21). In depleted plasma, the Control groups showed a 30% higher GPC4 content than ALS (Figure 5.21A), but with a higher variability as C1 showed a very faint band. This trend was in line with quantitative MS data (Table 5.8). On the other hand, neat plasma showed a much higher content in the ALS group (Figure 5.21B), more than three-fold, as obtained from plasma NCHs (Figure 5.14).

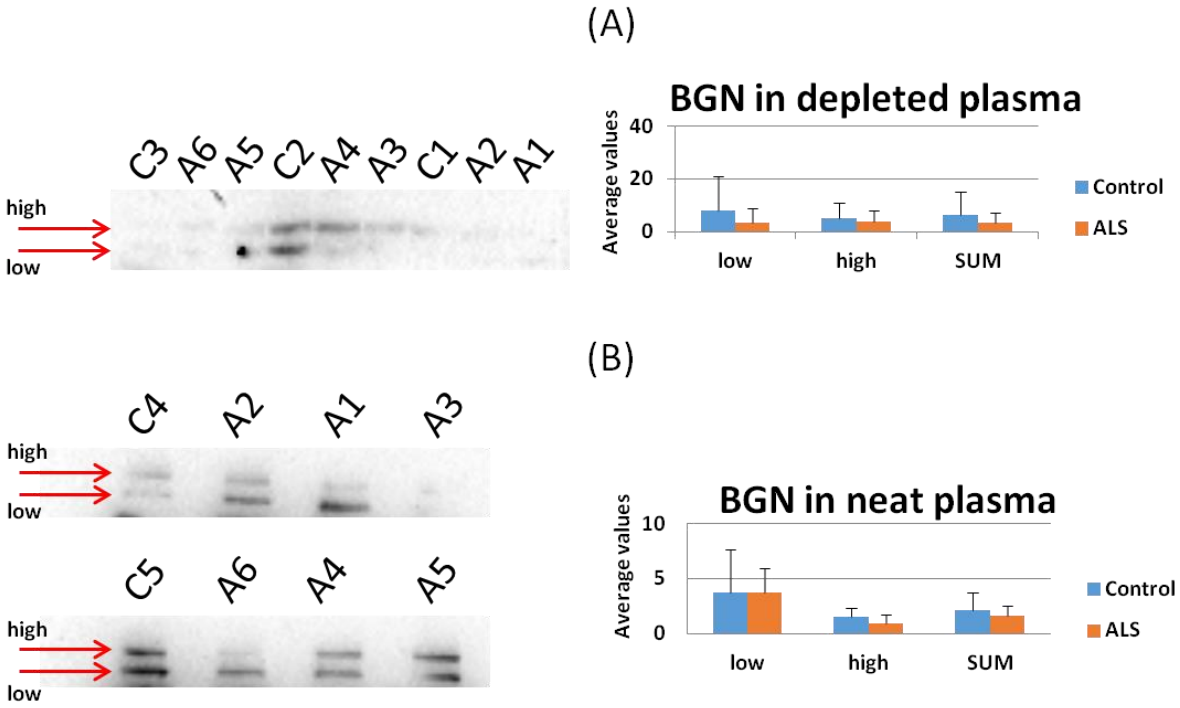


Figure 5.22. Immunodetection of BGN in depleted and neat plasma.

(A) and (B) show immunodetection and relative quantification of BGN in samples listed in Table 5.8. Quantification was performed normalising on C3 for (A) and C4 in (B).

Detection of BGN showed to be problematic in depleted plasma (Figure 5.22A), where four samples (C3, A6, A2 and A1) showed very faint immunoreactive bands, while in neat plasma BGN was easily detected in most of the samples (Figure 5.22B). In line with detection in plasma NCHs (Figure 5.17), two bands were detected for the same protein which presented a different trend of regulation with the Control group showing a higher expression of BGN than ALS (Figure 5.22). However, these data are in line with MS quantitative data (Table 5.8).

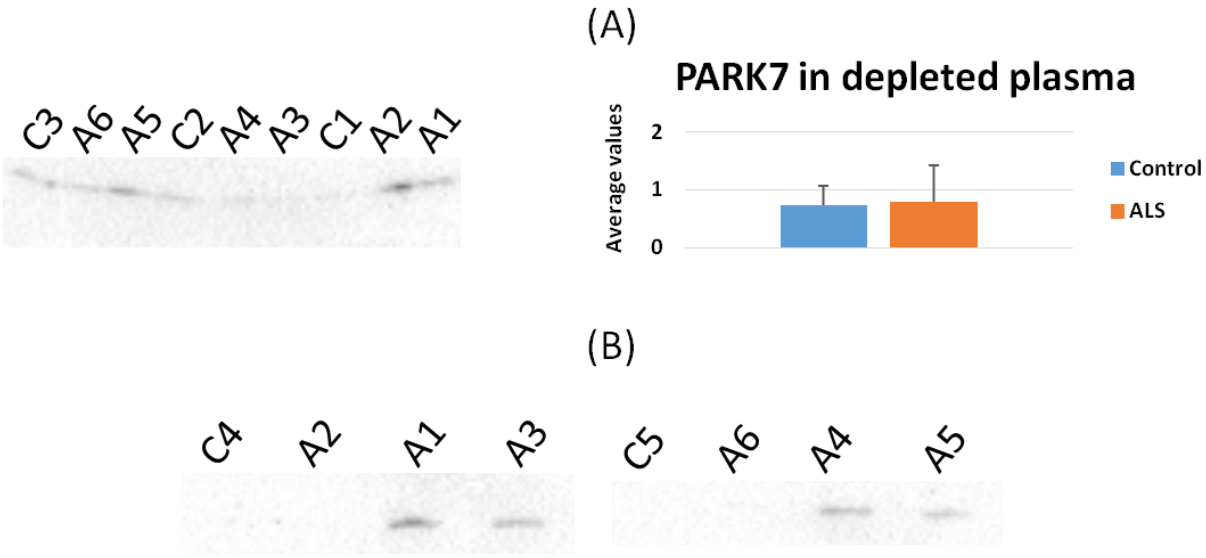


Figure 5.23. Immunodetection of PARK7 in depleted and neat plasma.
(A) and (B) show immunodetection of PARK7 in samples listed in Table 5.10. Quantification was possible only in depleted plasma (A) normalising each value on C3.

PARK7 was detected in all the depleted plasma samples, showing a slightly higher expression (about 10%) in the ALS group compared to Control (Figure 5.23A) while PARK7 was detected only in half of neat plasma samples (Figure 5.23B). Both Control neat plasma samples showed no PARK7 expression (Figure 5.23B). The same antibody detected two bands in NCHs (Figure 5.18), instead of one in neat plasma (Figure 5.23), suggesting that the nature of the sample can affect not only the overall detection but also selection of different species of the same protein.

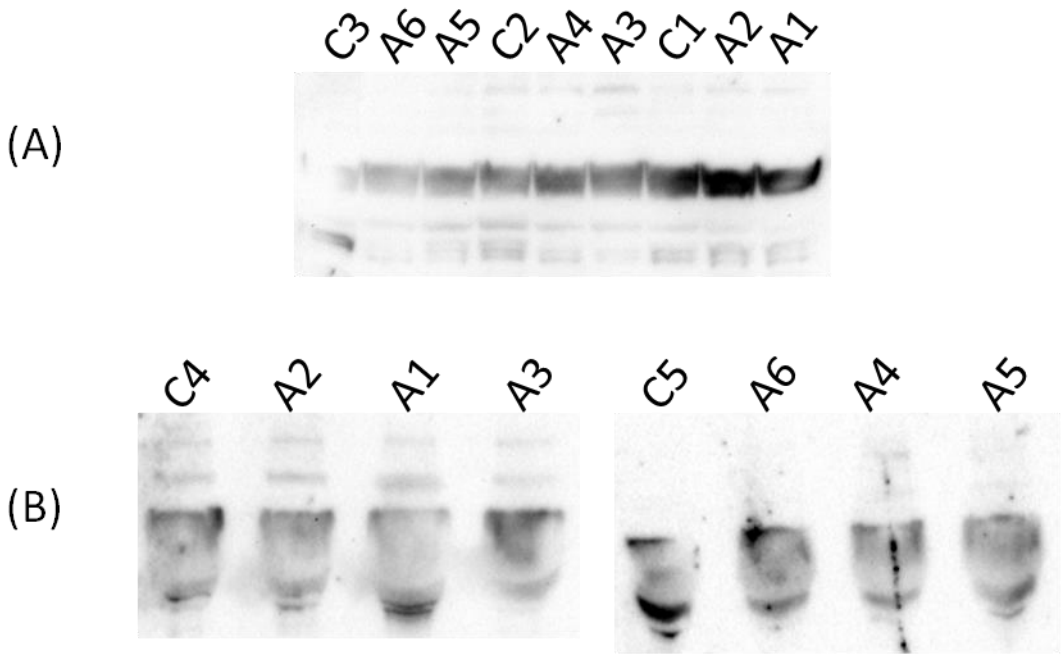


Figure 5.24. Immunodetection of SH3GLB2 in depleted and neat plasma.

(A) and (B) show immunodetection of SH3GLB2 (endophilin-B2) in the samples from the ALS patients and HC listed in Table 5.10. In both blots (A and B), Endophilin-B2 was detected in several bands and samples had different bands. In addition, depleted plasma (A) and neat plasma (B) showed different band profiles for SH3GLB2.

SH3GLB2 was detected in several bands (Figure 5.24), showing a different western blot profile from the one seen in plasma NCHs (Figure 5.15). Not all bands were present in all the samples, indicating also an inter-individual variability (Figure 5.24). The unclear pattern of the bands did not allow accurate quantification.

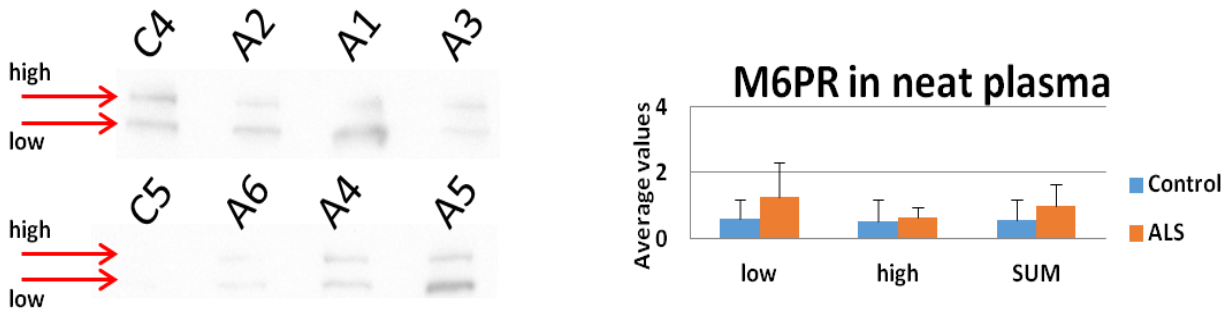


Figure 5.25. Immunodetection of M6PR in neat plasma.

The figure shows immunodetection of M6PR in samples from the ALS patients and HC listed in Table 5.10. Quantification was performed normalising each value on C4. Two bands were detected for M6PR, “low” and “high”, indicated by the arrows on the left-hand side of the blot.

M6PR could not be detected in depleted plasma in all samples tested. In addition, M6PR expression in neat plasma was minimal particularly in the HC group (Figure 5.25). The lower expression affected detection, but overall M6PR expression was 70% higher in the ALS group compared to HC (Figure 5.25). This result is in contrast to the trend of M6PR expression in plasma NCHs (Figure 5.19) and in line with MS data (Table 5.8).

5.9 Analysis of NCHs proteomic data without the calibrant channels

The TMTcalibrator workflow is designed to analyse in the analytical channels only those peptides that are expressed in the calibrant channels. Therefore, in our experimental setting this proteomic technique excludes the detection of those peptides that are likely to come from proteins that are typically present in plasma and that may be part of the NCHs under investigation (and not expressed in brain). For this reason and in order to provide a more comprehensive cover of all the molecules that assembles in the aggregates, we have looked in the datasets to uncover information related to NCHs proteins that have plasma origin.

The search performed after spectra acquisition with proteome discoverer 1.4 (PD) identified a total of 679471 PSMs, of which 323442 in TMTplex01 and 356029 in TMTplex02. 1.59% of the total PSMs identified (specifically 5147 in TMTplex01 and 5682 in TMTplex02) were not detected in any calibrant channels and in all analytical channels of the two TMTplexes. Taking into account only the PSMs belonging to single proteins (i.e. not shared among more than one protein group), 101 and 82 protein groups were identified in TMTplex01 and TMTplex02 respectively (Figure 5.26). Excluding three proteins only identified in the TMTplex01, the PSMs belonging to this set of proteins, with relative quantitative data, were not taken into consideration for quantitative analysis by the bioinformatic pipeline as they did not represent brain proteins. This has certainly affected the ultimate quantitative comparison of NCHs and their relative composition of plasma and brain-derived proteins.

Other proteins that could have the potential to represent good disease biomarkers in ALS are those included in plasma NCHs that are derived from plasma (i.e. not detected in brain calibrant) and that may be detected in only one of the two experimental groups (i.e. ALS or HC). However, no representative protein was detected from plasma NCHs, suggesting that the peptides in the calibrant

channels saturate signal acquisition hiding most of the non-calibrant derived peptides.

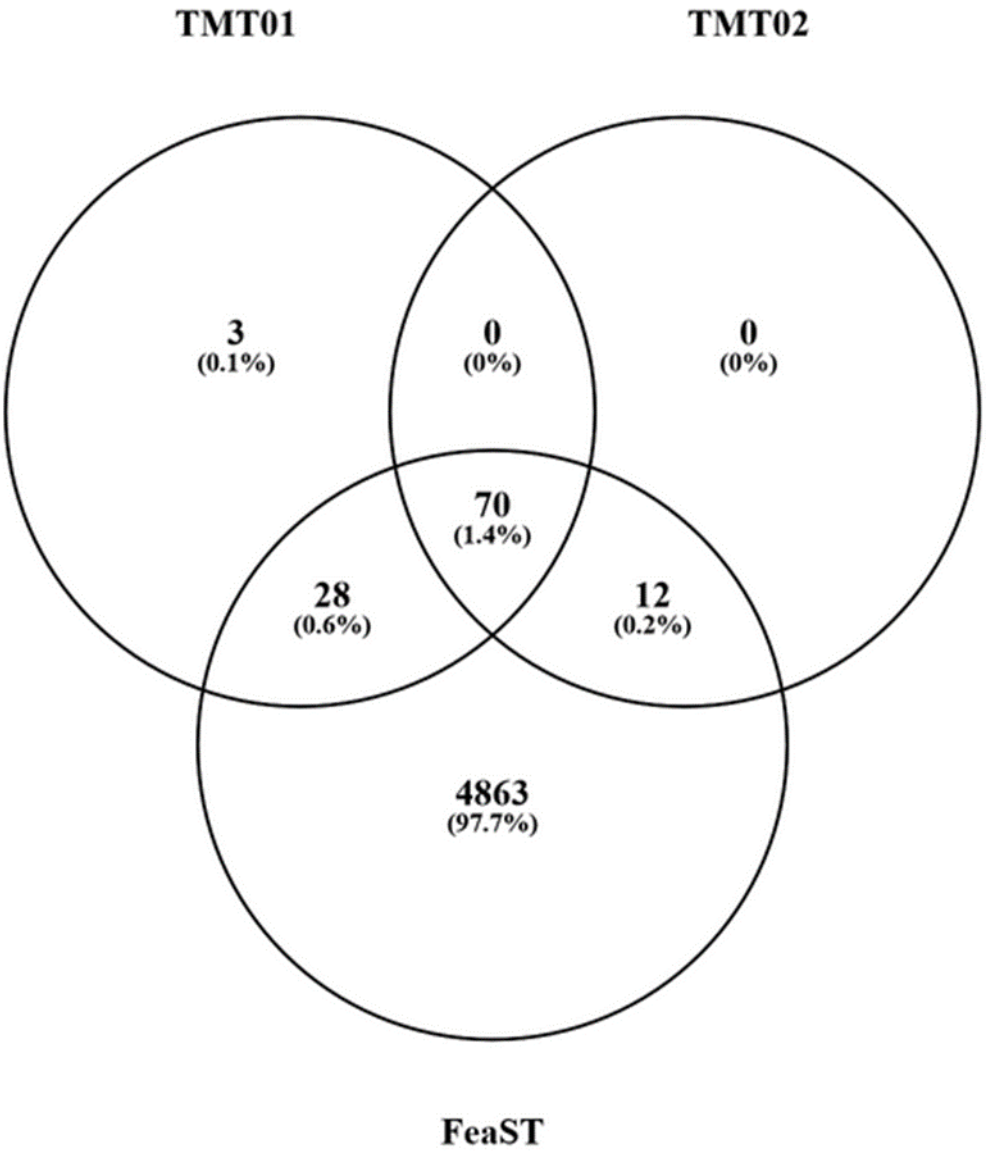


Figure 5.26 Proteins identified in the TMT 10plexes with no reference to the calibrant channels

Protein groups identified in TMTplex01 (TMT01) and in TMTplex02 (TMT02) derived from the PSMs identified in all the analytical channels and simoultaneously in no calibrant channels. The Venn diagram shows a comparison across the two TMTplexes and the protein lists generated by FeaST (FeaST). 70 proteins were shared among TMTplexes and Feast proteins, 28 between FeaST and TMT01, 12 between FeaST and TMT02, while no proteins were shared between TMTplex01 and TMTplex02.

5.9.1 Brain-derived exclusive proteins for each experimental group

Another information that could be useful in this experimental approach for biomarkers are the peptides, with relative proteins, shared between the calibrant (i.e. brain) and only one of the two experimental groups, here after called exclusive proteins. This could be investigated in the matrix generated by CalDIT before quantitative comparison performed by FeaST, here after called pre-FeaST matrix, where all the peptides identified in the Calibrant channels in both TMTplexes and semi-quantitative data from the analytical channels are included.

Pre-FeaST matrix contained 99750 peptides identified through PD search in both TMTplexes. Of these 99750 peptides 15290 (15.33%) were not identified in both ALS and HC channels, while 37638 peptides (37.73%) were identified in both experimental groups in at least four channels. This share of information was then analysed through FeaST and FAT, and the results have been already presented above in this chapter. These numbers suggest that some useful information for biomarkers study could have been filtered out. In fact, 40021 peptides (40.12%) were identified in at least 4 samples in ALS experimental group, and the 65.46% (26198 peptides) of these were identified in all 6 ALS samples (26.26% if considering the pre-FeaST matrix), while 42937 peptides (43.04%) identified in at least 4 HC samples, of which the 65.00% (27910) were identified in all HC samples (27.98% if considering the pre-FeaST matrix). Thus, there is still a small share of information containing semi-quantitative data of brain-derived peptides in plasma NCHs that is exclusive for ALS or HC experimental group (2.39% and 5.31% of the pre-FeaST matrix for ALS and HC, respectively).

As mentioned above, presence or absence of a specific peptide is given by the identification of at least four samples in both experimental groups. Thus, identification of a peptide in four samples from one experimental group and three sample from the other group represents the limit condition for the absence of that specific peptide in following FeaST analysis. At the same time this is the minimum condition that implies exclusive proteins for the two experimental groups, as it defines the identification of that peptide in one experimental group while the absence in the other one. However, the peptides identified with such a small difference in number of positive/negative samples for the identification of a peptide are very unlikely to be exclusive proteins. To increase likelihood of

characterising brain-derived proteins that are exclusive of the two conditions tested, ALS and healthy states, only the peptides identified in at least four samples in ALS or HC group and at the same time in none of the samples of the second experimental group were considered in the following analysis.

5.9.1.1 Exclusive proteins in ALS

In pre-FeaST matrix, 46 peptides were identified in at least four ALS samples and absent in all HC samples. Matching the protein groups identified through these peptides, with the protein list obtained from FeaST, five protein groups could be classified as ALS exclusive (Table 5.11).

Gene name	UniProt ID	Protein name	Number of peptides	Number of PSMs	Number of positive ALS samples
DCUN1D3	Q8IWE4	DCN1-like protein 3	1	2	4
NOTCH2	Q04721	Neurogenic locus notch homolog protein 2	1	4	4
PUS1	Q9Y606-2	Isoform 2 of tRNA pseudouridine synthase A, mitochondrial	1	2	4
CAMK2D	Q13557-4	Isoform Delta 4 of Calcium/calmodulin-dependent protein kinase type II subunit delta	1	1	4
PRKAA2	P54646	5-AMP-activated protein kinase catalytic subunit alpha-2	1	5	5

Table 5.11. Table with ALS exclusive proteins

Gene name: the recommended gene symbol used to officially represent a gene

Uniprot ID: Uniprot database protein identifier

Protein name: protein full name recommended by Uniprot

Number of peptides: number of peptides identified in at least four ALS samples and in zero HC samples

Number of PSMs: number of PSMs detected for each specific peptide

Number of ALS samples: number of ALS samples in which each peptide was detected

5.9.1.2 Exclusive proteins in HC

For HC experimental group, 169 peptides potentially belonging to exclusive proteins were identified. Matching the list of protein groups obtained from these 169 peptides with the list of protein obtained after FeaST, six protein groups were characterized (Table 5.12).

Gene name	UniProt ID	Protein name	Number of peptides	Number of PSMs	Number of positive HC samples
CEP170	Q5SW79-2	Isoform 2 of Centrosomal protein of 170 kDa	1	3	4
CSPG5	O95196-2	Isoform 2 of Chondroitin sulfate proteoglycan 5	1	2	4
SCN8A	Q9UQD0-5	Isoform 5 of Sodium channel protein type 8 subunit alpha	1	2	4
MRPS34	P82930	28S ribosomal protein S34, mitochondrial	1	2	4
RBKS	Q9H477	Ribokinase	1	2	4
ANKRD35	Q8N283	Ankyrin repeat domain-containing protein 35	1	1	4

Table 5.12. Table with HC exclusive proteins

Gene name: the recommended gene symbol used to officially represent a gene

Uniprot ID: Uniprot database protein identifier

Protein name: protein full name recommended by Uniprot

Number of peptides: number of peptides identified in at least four HC samples and in zero ALS samples

Number of PSMs: number of PSMs detected for each specific peptide

Number of HC samples: number of HC samples in which each peptide was detected

5.10 Conclusion

The use of TMT calibrator proteomics described in this chapter has been highly effective and ultimately has succeeded in identifying a significant number (almost 5000) of brain-derived proteins in plasma NCHs, while increasing exponentially the detection of low abundance proteins like neurofilaments (Nf) in the mix of proteins normally populating a complex matrix like plasma, both in its soluble and aggregated state. We consider this a significant achievement and a major step forward in the use of proteomics to unravel the complex molecular framework of biofluids in search of biochemical markers of disease. MS proteomics is normally limited by the interference of high and relatively high abundance proteins that may also have relevance as biomarkers, but that ultimately limit the process of identification of less represented proteins coming from the brain. Together with neurofilaments (Nf), the list of ALS-relevant proteins identified using the ALS-related gene list database (Table 5.5) which have been detected in our brain/plasma NCHs experiment are also normally escaping detection by different standard proteomic techniques in virtue of their low-abundance in biofluids.

The NfH isoform in proteins aggregates showed almost statistically significant regulation ($p=0.06$) when comparing ALS and HC NCHs adding to previously reported data in plasma which support the use of this protein as a marker of neurodegeneration. The confirmation that proteins like Nf can be packaged into circulating aggregates obtained by TMT proteomics as well as using gel separation and TEM should raise important question on the importance of the state of relevant biomarkers (i.e. soluble vs aggregated) when they are the target of novel diagnostic tools based on immunodetection.

Another known and important protein detected in this study is FUS, which showed to be expressed differently between ALS and HC groups in NCHs. FUS is known to be a key protein ALS protein; FUS gene mutations are causative of ALS and these genetic defects lead to deranged protein translation and and altered cell

homeostasis (253–256). However, FUS was not linked to any regulated pathway obtained from our FAT functional analysis.

Those proteins detected by MS which were not processed through the TMTcalibrator™ workflow have been assessed as potential biomarker candidates looking at whether they were uniquely present or absent in ALS. However, few of these proteins were singled out and all were identified by one unique peptide and maximum five PSMs, not providing strong evidences at MS level for further development of these proteins as biomarkers. The absence of representative plasma proteins within the NCHs must be attributed to the TMTcalibrator™ setting, where the calibrant represents more than the half of the entire amount of proteins loaded in each TMTplex. In this way, the signal from brain-derived peptides (i.e. calibrant channels) prevails over non-brain-derived peptides, especially in the data dependent acquisition setting used in this workflow. Hence, equally informative signals coming from plasma may be lost as the platform privileges the detection of signals of brain origin.

Functional analysis of the TMT calibrator data has shown features which are in keeping with the ALS pathobiology, confirming the reliability of the approach adopted in this study. In addition, it was possible to find biochemical pathways that had not been previously linked to ALS, suggesting new research avenues in this field and the importance of NCHs as conduits of pathology and testbeds to discover novel biomarkers and therapeutic leads. Increasing stringency of our bioinformatic analysis, we have been able to generate a list of potential biomarker candidates for ALS to be considered for future validation studies. These are brain proteins in circulation found in systemic aggregates, whose regulation is in relation to the pathological state. The assessment of these clusters of proteins allows us to appreciate functional and biochemical changes with a high degree of affinity to known pathogenic mechanisms already described in ALS and to novel molecular pathways. Future work will have to include the selection of the best candidates to be considered for further validation in larges cohorts of ALS and HC, using a different and high-throughput methodological approach like immunodetection.

In this study, the involvement of glycosaminoglycan metabolism, lysosome, synthesis of phosphatidic acid and synaptic vesicle pathway was investigated by

immunodetection of FMOD, GPC4, BGN, M6PR, SH3GLB2 and PARK7. Unexpectedly, the western blot analysis in plasma NCHs did not confirm the same regulation showed by proteomics analysis in largely the same individuals evaluated by TMTcalibrator™. MS and antibody-based techniques have different analytical sensitivity and specificity. Protein quantification by MS relies first of all on the number of PSMs and peptides identified in the complex mix of proteins subjected to analysis. A number or most of peptides belonging to target proteins may not be detected, possibly because of the low peptide abundance in respect to other co-eluting peptides or for the presence of PTMs not included in the search step. Our effort to extend the validation by immunodetection to total plasma and depleted plasma samples was useful to study the potential to test these molecules in a technically less challenging method, ultimately without the need for NCHs extraction. The differences in protein expression in the study groups seen using MS and immunodetection could be justified by the different samples used in the western blot and MS analysis and by the biology of the protein targets which condition their detection using different methods. However, PARK7 showed presence exclusively in ALS plasma samples. This would suggest further validation steps by increasing the number of samples and arrive to an accurate estimate of the potential of this protein candidate as biomarker.

In conclusion, apart from those proteins which control lipoproteins homeostasis, the pathways we have singled out using a pioneering tissue/fluid proteomics do not seem to have been previously linked to the pathogenesis of ALS, suggesting novel potential routes for investigation into neurodegeneration and ALS. It is also fair to mention that most of the data and literature available is limited to microorganisms and cell models, while there is almost no data on blood circulating protein aggregates, their composition and putative role in the organism homeostasis and therefore, it is important to remain cautious in any speculation on the relevance of our data. Future studies will have to be undertaken using cell and/or animal models to establish whether the molecular pathways we have identified in our Brain/plasma NCHs proteomic studies are truly implicated in the pathogenesis of ALS or are part of other pathogenic mechanisms more broadly implicated in neurodegeneration.

6 Altered biochemical properties in neurofilament-containing heteroaggregates (NCHs)

6.1 Introduction and aim

In consideration of the presence of protein aggregates which have been linked to the seeding behavior of disordered protein (e.g. oxidized proteins), it has been proposed that the pathogenesis of Amyotrophic Lateral Sclerosis (ALS) may have features in common with other rapidly developing neurodegenerative diseases like Prion's disease (80,144,145,257). The data reported in Chapter 4 showed a degree of protein resistance to digestion with trypsin of these formations (Figure 4.19). These preliminary data require further investigations to ascertain whether NCHs may truly present some of the altered biochemical properties seen in prion disease.

To shed some light on this possibility, plasma NCH samples obtained from individual patients affected by ALS along with healthy controls (HC) are examined by transmission electron microscopy (TEM) and tested for resistance to digestion with a number of proteases. This Chapter will be divided into two parts. The first part will show new data collected by TEM, while in the second section we will describe NCHs resistance to digestion using Trypsin, Chymotrypsin, Enterokinase and Calpain. The effect of the same proteases (excluding trypsin the effect of which has already been tested in Paragraph 4.6) will also be assayed in brain NCHs to shed light to the biochemical behavior of NCHs coming from different tissues. This new set of experiment has the purpose of testing the prion-like disease hypothesis in the field of neurodegenerative diseases studies and in the context of ALS.

6.2 Transmission electron microscopy (TEM) imaging on NCHs

To evaluate possible differences in the macromolecular structures present in plasma NCHs from ALS and healthy individuals, I have analysed by TEM three HC samples (HC4, HC5 and HC6 listed in Table 6.1) and three ALS samples (ALS1*, ALS3* and ALS5* listed in Table 6.1).

Sample ID	Gender	Ethnicity	El-Escorial	Clinical Onset	Age at visit (year)	ALSFRS-R	Rate of progression at visit
ALS1*	F	Caucasian	Definite ALS	LL	61,11 - 61,99	19 - 36	0,649 - 0,997
ALS2*	F	Caucasian	Definite ALS	UL	68,45 - 68,80	35	0,821 - 1,111
ALS3*	M	Caucasian	Definite ALS	LL	59,62 - 61,18	35 - 38	0,436 - 0,853
ALS4*	M	Caucasian	Definite ALS	UL	64,40 - 65,82	8 - 12	1,423 - 2,369
ALS5*	M	Caucasian	Definite ALS	LL	63,05 - 64,79	22 - 37	0,623 - 0,801
HC1	F	Caucasian	NA	NA	62.48	NA	NA
HC2	F	Caucasian	NA	NA	67.50	NA	NA
HC3	M	Caucasian	NA	NA	60.64	NA	NA
HC4	M	Caucasian	NA	NA	68.28	NA	NA
HC5	M	Caucasian	NA	NA	62.54	NA	NA
HC6	F	Caucasian	NA	NA	63.53	NA	NA

Table 6.1 List of samples selected for further transmission electron microscopy analysis (TEM) and digestion by proteases.

In light orange background the Amyotrophic Lateral Sclerosis (ALS) patients and in light blue background the healthy control (HC). If the enriched fractions were obtained using more than one time point plasma sample for a single patient, the range values for age at visit, ALSFRS-R and progression rate were reported. HC3 and HC6 are the same samples listed in Table 5.1; *: samples obtained from the same ALS patients listed in Table 5.1, but at different time points.

Sample ID: ALS= ALS patients; HC= Healthy Control

Gender: M= Male; F= Female

El-Escorial: diagnostic classification of ALS diagnosis (88)

Clinical Onset (site of initial clinical signs): LL= Lower Limbs; UL= Upper Limbs

Age at sampling: age of the patient at blood sampling

ALSFRS-R (ALS Functional Rating Scale revised): level of neurological impairment across different clinical domains (1-48, with higher neurological impairment with lower values)

Rate of progression at sampling: calculated as 48 - ALSFRS-R score at sampling time/disease duration from onset of symptoms to sampling time

Analysis of these samples confirmed the presence of amorphous electron-dense NCHs (Figure 6.1), including round particles of different size in both HC and ALS samples (Figure 6.2).

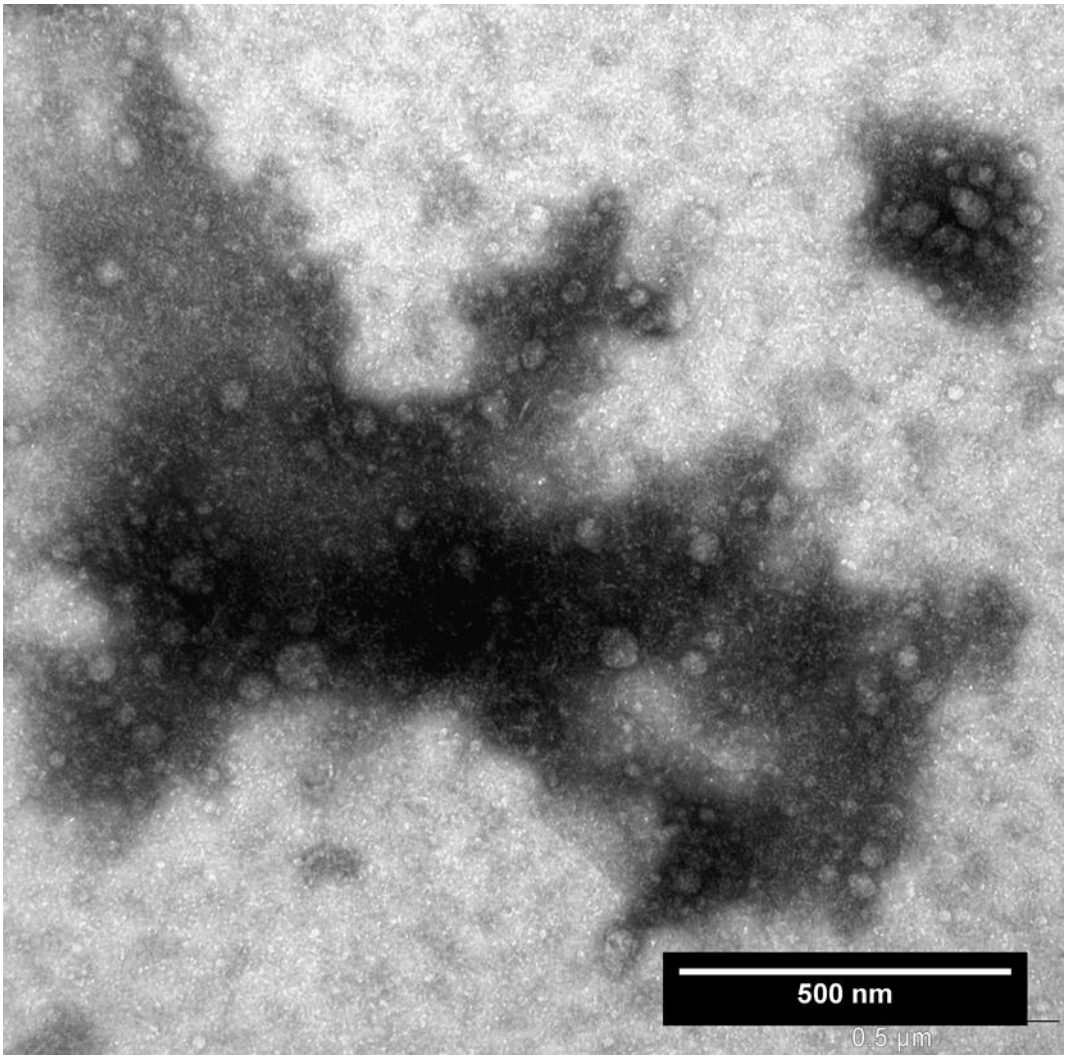


Figure 6.1. NCH amorphous globular macrostructures.

A representative micrograph showing an amorphous globular formation in aggregates enriched fractions. This image was captured from the grid loaded with HC6 sample.

TEM analysis of the NCHs fractions from HC uncovered the presence of round small formations within and outside the body of the bigger globular electron-dense formations representing the NCHs under investigation. These particles may be micelles as reported in the literature. In the context of our experiments, it is likely that they may be composed of lipids, detergents or lipoproteins (167,258). Due to the nature of the samples and the enrichment process none of these options could be discarded. However, we did not detect the same particles in the brain samples analysed in Paragraph 4.4.1 despite the amount of lipids in brain

tissue and the fact that the same aggregates enrichment process was applied. We could speculate that these particles are in fact made of lipoproteins, such as very low-density lipoproteins (VLDL), low-density lipoproteins (LDL) and high-density lipoproteins (HDL) (Figure 6.2). In support of this hypothesis, lipoproteins were also detected by mass spectrometry (MS) and biochemical pathways involving these proteins were significantly regulated in the proteomic study of the NCHs fraction (Table 5.6).

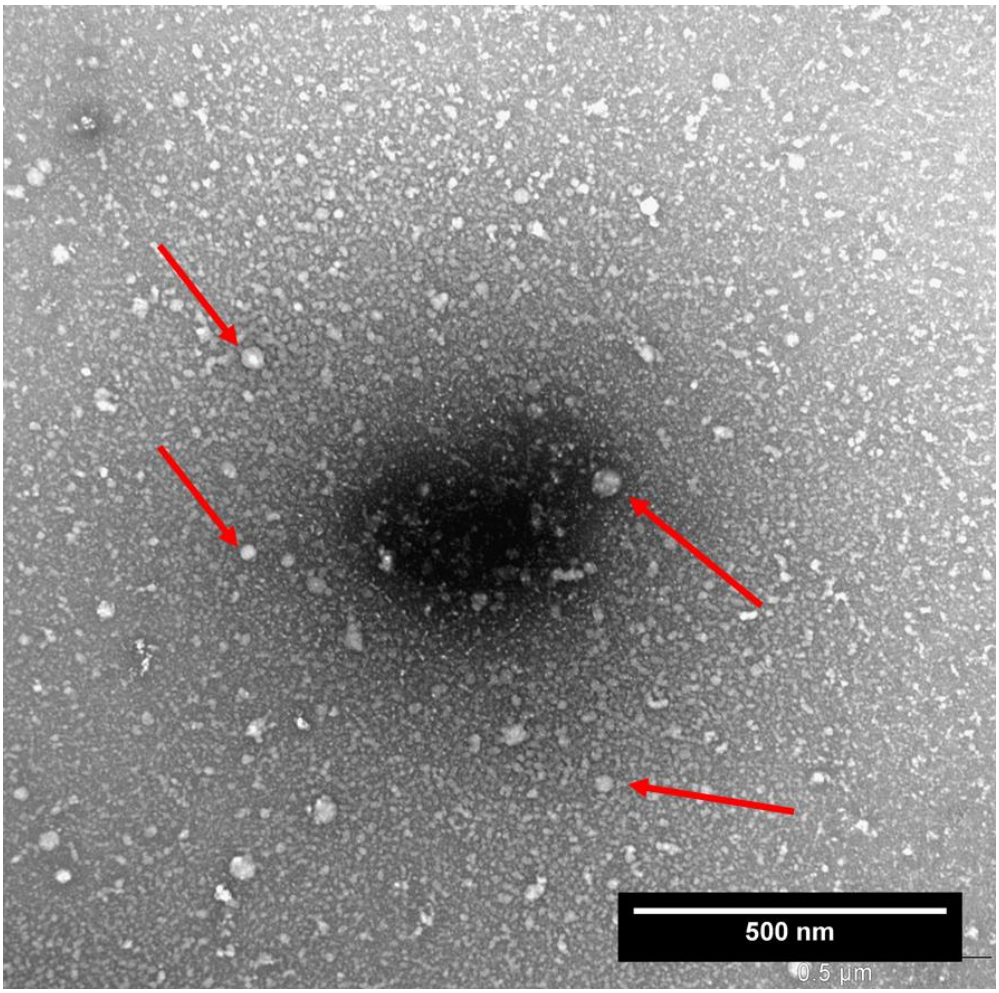


Figure 6.2. Particles identified in the aggregates enriched fractions.
A representative micrograph showing presence of particles (or micelles) in aggregates enriched fractions. The image is captured from the grid loaded with the ALS5* sample and show an electron-dense area surrounded with micelles of different sizes (highlighted by the red arrows).

Inspection of the grids prepared with NCHs enriched from ALS patients, showed the presence of small filamentous fragments with a rough surface (Figure 6.3) that were not detected in HC grids. Although this is far from being certain and difficult to prove in the context of our experiment, these fragments visually

resemble the formations observed in brain enriched fractions shown in Figure 4.10 .

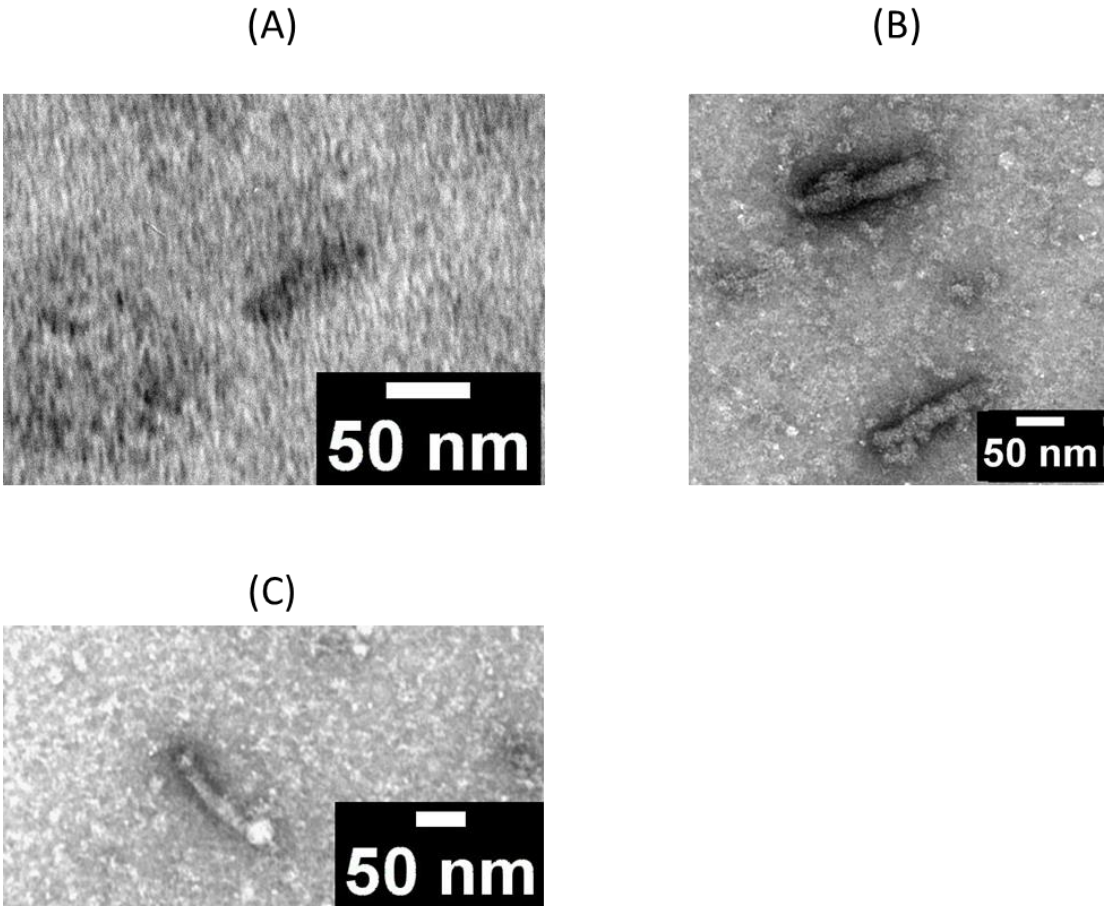


Figure 6.3. Details of particles detected in TEM micrographs of ALS NCHs.
((A) ALS1* grid with a 13 nm thick and 70 nm long fragment. (B) ALS3* grid with two fragments both with a 20 nm thickness. The lower fragment is 120 nm long and while the upper fragment is 145 nm long. (C) ALS5* grid with a 16.5 nm thick and 130 nm long filament.

This type of particles was found only in the ALS samples and they were similar although shorter to those found in ALS Brain NCHs (recalling prion-like structures), albeit slightly thicker. To date, there is no evidence in the literature of circulating particles in blood stream of ALS patients and animal models, with an appearance recalling prion proteins. Further investigation on the nature of these particles will require 1) a more extensive comparison with brain aggregates, 2) a bigger number of samples and 3) comparative studies with prion particles obtained from positive control samples (e.g. Creutzfeldt-Jakob disease).

6.3 NCHs resistance to digestion

One of the main biochemical characteristics of proteins behaving like prions is their resistance to digestion with proteases (146). We have already shown preliminary data on plasma and brain NCHs in Chapter 4, where ALS and HC plasma NCHs and ALS brain NCHs samples tested showed a degree of resistance to trypsin digestion (digested samples appeared like a light smear in the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 4.16). In the same experimental setting, neither neurofilament heavy (NfH) (Figure 4.17) nor transactive DNA binding protein (TDP43) (Figure 4.18) showed resistance to trypsin digestion, while a small portion of ubiquitinated proteins was still detectable after digestion of brain aggregates (Figure 4.19).

To further investigate resistance to protease digestion, NCHs enriched from the ALS and HCs plasma samples listed in Table 6.1 were treated with trypsin, chymotrypsin, enterokinase and calpain, as described in Paragraph 2.6. The digestion patterns were evaluated by gel electrophoresis followed by zinc staining to highlight the general digestion profiles and by NfH immunodetection to test any resistance to protein digestion of this specific protein.

All NCHs samples from ALS patients and HCs listed in Table 6.1 were ran after digestion with different proteases along with the undigested matched sample (Figure 6.4) and subjected to zinc staining. In total, six gels were ran to evaluate digested and undigested NCHs from ALS and HCs (including one ALS and one HC sample in each gel). The aim was to identify gel profiles indicative of (partial or total) resistance to protease digestion. For internal normalization in the quantitative analysis of the digestion profiles, we have used the gel marker as described in Paragraph 2.6.1. To test NfH protease resistance in the same NCHs, NfH immunodetection was performed in the same gels in digested and undigested samples.

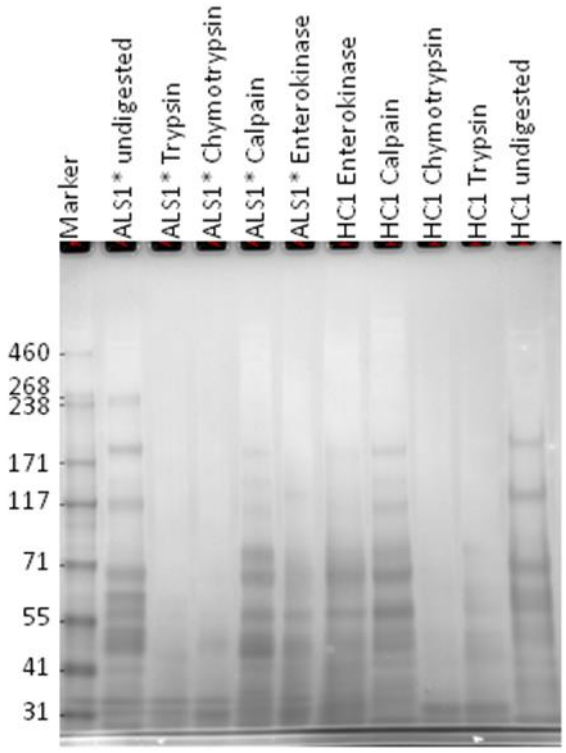


Figure 6.4. SDS-PAGE of ALS and HC NCHs samples before and after digestion with proteases.

The figure shows an example of gel electrophoresis and zinc staining used for the analysis of protease digestion in pairs of ALS and HC samples. This gel is reported as ALS1_HC1 and shows products of trypsin chymotrypsin, calpain and enterokinase digestions generated from ALS1* and HC1 (along with the same undigested samples as reference).

4.03 µg of total proteins were loaded in each lane (samples ALS1*-5* and HC2-6) along with 10 µl of marker in each gel. For HC1 samples, 2.23 µg were loaded in each lane because of the low protein concentration in these samples. The coefficient of variation (CV%) of the total intensity calculated on the undigested lane for each individual did not indicate equal loading for all samples (Table 6.2). Because unequal loading may have affected the final results, a correction factor was generated for each sample in order to reduce the bias related to different protein loading. Correction factors (CFs) were calculated as the ratio between the total intensities of each undigested sample (i.e. sum of all the bands) and the marker total intensities from the same gel (Table 6.3). For the quantification of each band from digested and undigested samples, each band intensity was first normalised with the closest band of the gel marker and then further normalised with the CFs specifically calculated for each sample in the experiment.

Because NCHs samples generated from HC1 had a low protein concentration, the amount of total protein loaded was different from the other ALS and HC samples and required an additional loading factor (LF) to be calculated before protease digestions data analysis in order to have an even measurement of intensities. Hence, the maximum possible amount was loaded for each HC1 sample, especially to increase peptides detection for the digested samples where most of the proteins are supposed to be hydrolysed, and LFs calculated as the ratio between the µg loaded for each sample undergoing digestion by trypsin, chymotrypsin, calpain or enterokinase and the related HC1 undigested sample (Table 6.4). Each LF was then applied together with the HC1 CF (Table 6.4) for the final analysis.

Total intensity for:	Gel1 (ALS1_HC1)	Gel2 (ALS2_HC2)	Gel3 (ALS3_HC3)	Gel4 (ALS4_HC4)	Gel5 (ALC5_HC5)	Gel6 (HC6)	Mean	St. Dev.	CV%
Marker	107704	109638	102576	114133	100891	95666	105102	6664	6.3
ALS	62481	57628	58192	55143	46977	-	56084	5735	10.2
HC	40800	59041	39941	46063	40622	50107	46095	7477	16.2

Table 6.2. Marker, undigested ALS and HC sample intensities.

The total bands intensity is obtained as the sum of the intensities of all bands in a lane after the images were processed by ImageJ, and it is shown for the the gel marker, for ALS and HC undigested samples in each gel (Gel1-6). Mean, standard deviation (St. Dev.) and coefficient of variation (CV%) are shown for each group. The low CV% of the marker confirmed equal loading which could be used as reliable reference signal. A higher CV% was calculated for ALS and in particular for HC samples, despite attempts to obtain equal loading. In grey colour-code the Mean, St. Dev. and CV% calculated for the total intensities.

	ALS1*	ALS2*	ALS3*	ALS4*	ALS5*	HC1	HC2	HC3	HC4	HC5	HC6
Correction Factors	0.58	0.56	0.57	0.48	0.47	0.38	0.54	0.39	0.40	0.40	0.52

Table 6.3. Correction factors (CFs) for each ALS and HCs sample calculated using undigested samples and marker total intensities.

The correction factors (CFs) were applied to quantify differences in NCHs profiles obtained by SDS-PAGE from undigested and digested samples. Each band under analysis, was first normalized with the marker band closer to the molecular weight (MW) of the bands under analysis and then divided by the sample CF. ALS1*-5* and HC1-6 indicates the samples used in this experiment with their specific CF.

HC1 samples					
	Undigested	**Trypsin**	**Chymotrypsin**	**Calpain**	**Enterokinase**
Sample loaded (µg)	2.23	3.27	3.32	3.34	3.45
Loading Factor	1	1,466837	1,492731	1,498645	1,550701
Correction Factor	0,3788	0,5557	0,5655	0,5677	0,5874

Table 6.4. Loading and correction factors for the HC1 samples.

To adjust for the uneven loading of HC1 samples, loading factors (LFs) were calculated to correct intensity values and allow statistical analysis. The CFs (Table 6.3) were adjusted for the different amount of proteins loaded for each HC1 sample with the relative LFs.

6.3.1 Undigested samples

The gel profiles of the undigested NCHs fractions from ALS patients and from HCs enriched from plasma were evaluated taking into account the adjustments for unequal samples loading reported above (Figure 6.5).

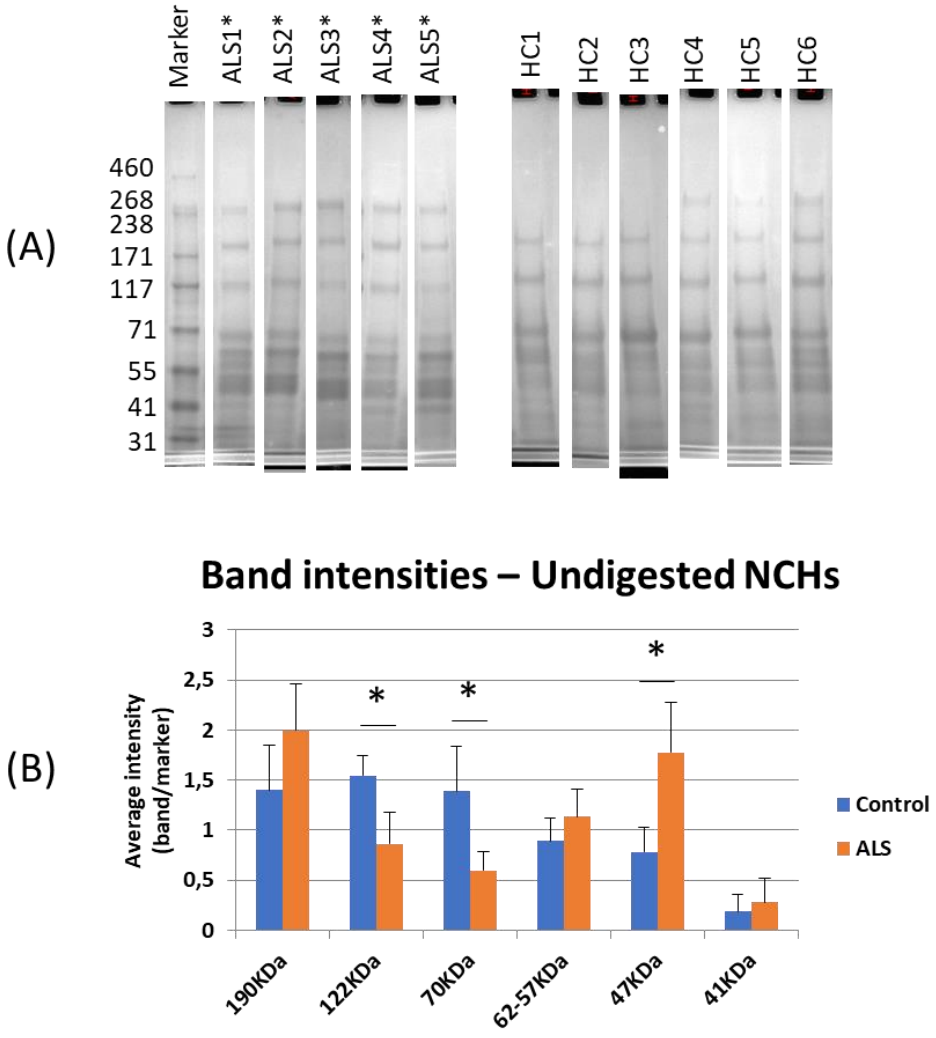


Figure 6.5. Undigested plasma ALS and HC NCHs profile.

(A) Gel profile after zinc staining of undigested plasma NCHs samples including ALS (ALS1*-5*) and HC (HC1-6). (B) Quantitative analysis of the band profile comparing ALS and HCs. Here bands at 190 KDa molecular weight (MW) were normalised with the closest Marker band at 171 KDa, the 122 KDa bands with the Marker band at 117 KDa, the 70 KDa bands with the Marker band at 71 KDa, the 62-57 KDa range bands with the Marker band at 55 KDa, the 47 KDa bands with the Marker band at 55 KDa and the 41 KDa bands with the Marker band at 41 KDa. *= p-Value < 0.05.

The profile of the undigested plasma NCHs from ALS and HC samples shared bands at the following molecular weight (MW): 190, 122, 70, 47 and 41 KDa as well as an area between 62 and 57 KDa (in this MW range, it was not possible to define a single band in all the samples under analysis) (Figure 6.5A). Plasma ALS NCHs also showed a band at 268/238 KDa, but this was not taken into consideration because three out of six HC samples did not display the same band (Figure 6.5A). Quantitative analysis of these bands between the two groups and statistical analysis showed that bands at 122 and 70KDa were more intense in the HC NCHs group (p= 0.004 and 0.006 respectively), while the band at 47 KDa was more intense in the ALS NCHs group (p= 0.006) (Figure 6.5B).

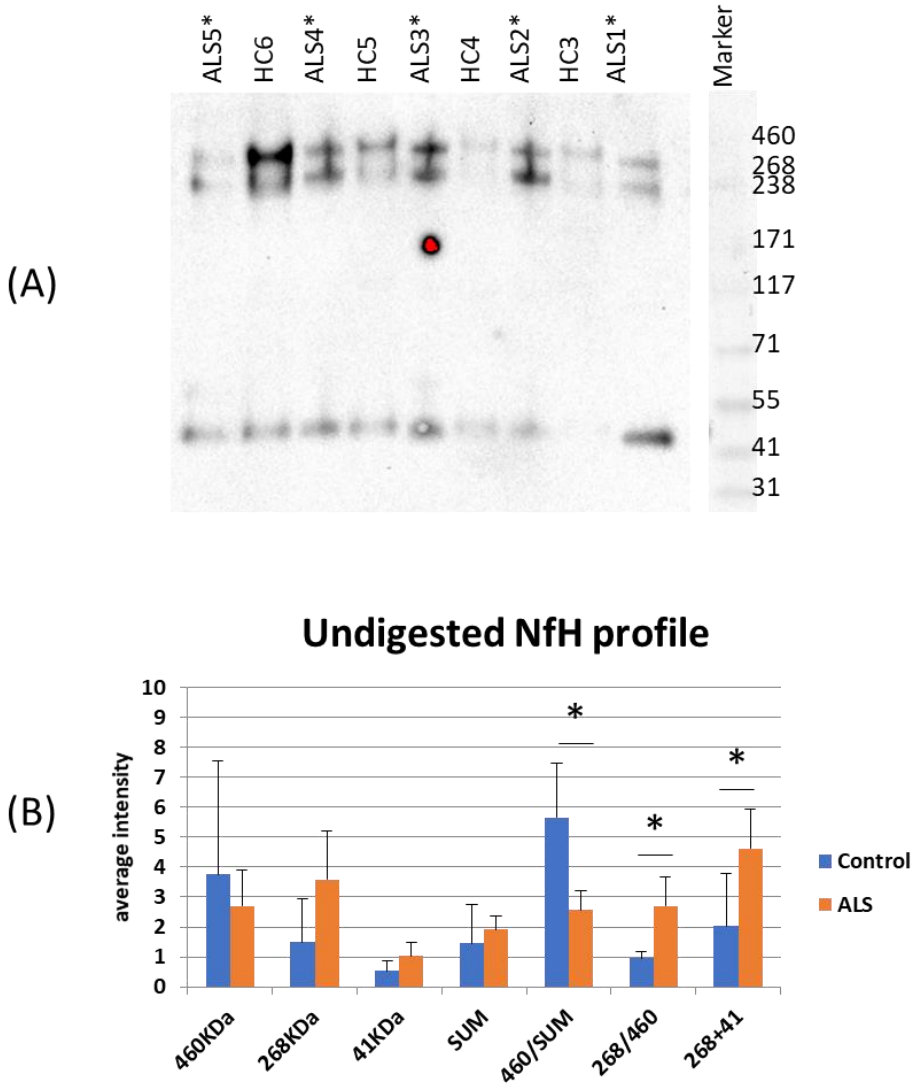


Figure 6.6. NfH profile in undigested plasma NCHs.

(A) Western blot against NfH revealed bands at 460 KDa, 268 KDa and 41 KDa. (B) Quantitative analysis of the bands detected, including the sum of the three bands detected (SUM), the ratio between the 460 KDa band the SUM (460/SUM), the ratio between the 268 and 460 KDa bands (268/460) and the sum of 268 and 41 KDa bands (268+41). Band densities were normalised on ALS1*. *= p-Value < 0.05.

Western blot against NfH in the undigested plasma NCHs revealed three bands at 460, 268 and 41 KDa (Figure 6.6A). Quantitative analysis of the NfH bands in ALS and HC samples showed that the overall NfH intensity (the sum of the three NfH bands: SUM) was higher in the ALS samples compared to HC samples although not statistically significant (p= 0.52; Figure 6.6B), while the ratio between the intensity of the 460 KDa band and the overall NfH intensity (460/SUM), the ratio between the intensities of the 268 and the 460 bands (268/460) and the sum of the bands at 268 and 41 KDa (268+41) showed a statistically significant higher expression in ALS compared to HCs (p= 0.032, 0.018 and 0.049 respectively; Figure 6.6B). These data are in line with previous NfH quantification in plasma NCHs. It is possible to postulate that the relative quantification of different NfH species (fragments) in plasma NCHs may be more informative than other more general NfH quantification.

6.3.2 Trypsin digestion

Trypsin is a serine protease, which cuts peptide chains at lysine or arginine. This proteolysis has been previously reported to be ineffective in prion digestion (146). The same enzyme was extensively used in sample preparation for our MS-based proteomics experiments that makes up the work presented in this thesis. Therefore, the analysis of NCHs resistance to protein digestion can help in defining possible shortfall in MS analysis of circulating aggregates.

The plasma NCHs profiles after trypsin digestion showed few bands ranging from 71 KDa (ALS3*, HC1, HC2 and HC6) to a lower molecular weight (Figure 6.7A). There was a high sample-to-sample variability with regards to the detected bands. The HC group showed a slightly higher degree of resistance (Figure 6.7A). However, the only band shared by all ALS and HC samples at 45 KDa was slightly more intense in ALS samples (Figure 6.7B).

NfH detection after digestion showed similar results (Figure 6.8). Most of the samples did not show any NfH band, while ALS5*, ALS3* and HC5 showed a

clear NfH positive band around 41 KDa (Figure 6.8). The overall results may indicate a degree of NCHs resistance to trypsin digestion, supporting the need to consider alternative proteases for future NCHs analysis by MS-based bottom up proteomics (Figure 6.7A). Also, NfH in an aggregated-state showed a certain degree of resistance which was certainly more pronounced in ALS samples. This biochemical change could have had a potential effect in this protein quantification in the reported MS experiments (Figure 6.8). In our in TMTcalibrator™ proteomic study, NfH was (almost significantly) up-regulated in ALS NCHs compared to HC (Paragraph 5.5). This observation and the incomplete digestion by trypsin would support further studies into the role of this Nf isoform in the pathogenesis of ALS.

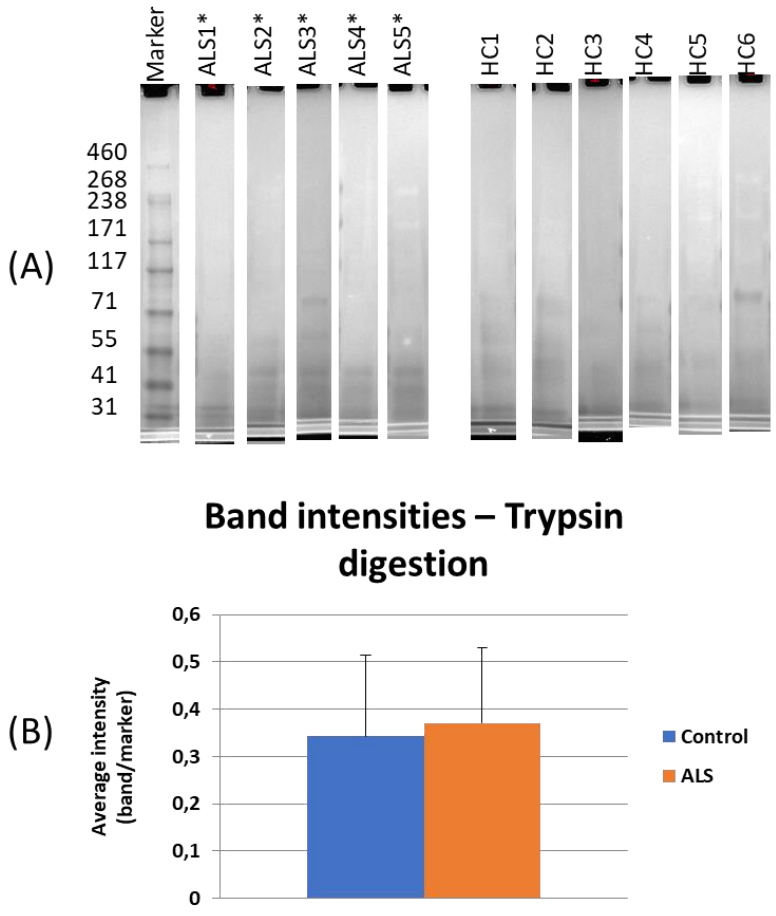


Figure 6.7. Plasma NCHs profile after trypsin digestion.

(A) Plasma NCHs from ALS (ALS1*-5*) and HC (HC1-6) digested with trypsin. (B) Quantitative analysis of the common band detected at 45 KDa and normalised with the adjacent Marker band at 41KDa.

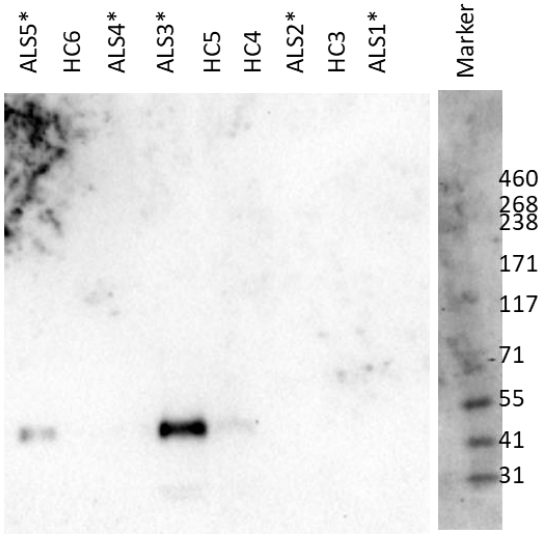


Figure 6.8. NfH profile in plasma NCHs digested with trypsin.

Western blot against NfH revealing a band around 41 KDa for ALS5*, ALS3* and a much fainter band at the same MW for the HC5 sample, while all other samples show a complete digestion. These findings indicate a degree of resistance to digestion with trypsin for NfH included in NCHs.

6.3.3 Chymotrypsin digestion

Chymotrypsin is a serine protease that hydrolyzes peptide bonds with aromatic or large hydrophobic side chains (tyrosine, tryptophan, phenylalanine, methionine and leucine). This protease was tested because it is included in the same family of trypsin, and both proteases were detected in proximity of Nf aggregates in motor neuron axons in spinal cord from ALS patients (259).

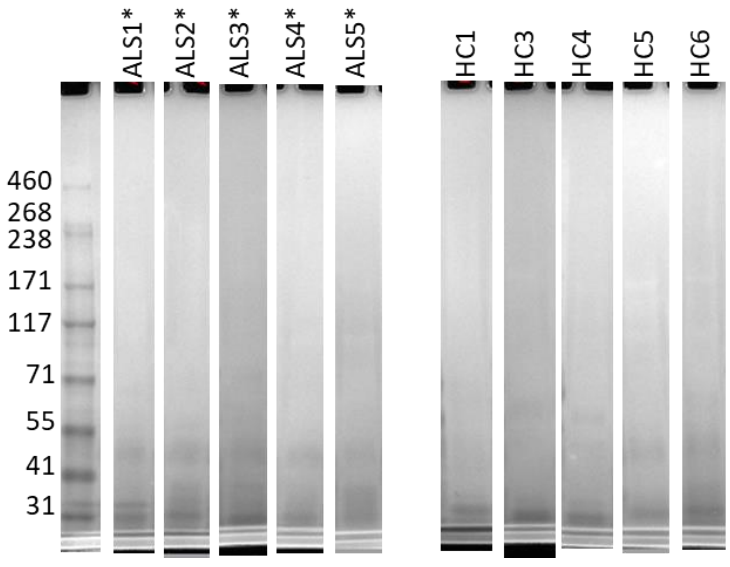


Figure 6.9. Plasma NCHs profile after chymotrypsin digestion.

Plasma NCHs samples from ALS (ALS1*-5*) and HC (HC1, HC3-6) digested with chymotrypsin. This digestion profile is represented by a few and faint bands, suggestive of a very low degree of resistance to digestion. Due to the low intensity of the bands and the background it was not possible to quantitate and compare the two experimental groups. HC2 was not included in the experiment as there was not enough NCH material for the digestion.

Plasma NCHs showed a very low or no resistance to digestion with chymotrypsin (Figure 6.9). ALS NCHs samples showed a very faint band around 41 KDa that was not visible in HC sample, while both experimental groups showed the same band below 31 KDa (Figure 6.9). However, due to low band intensity and staining background, it was not possible to perform a reliable quantification.

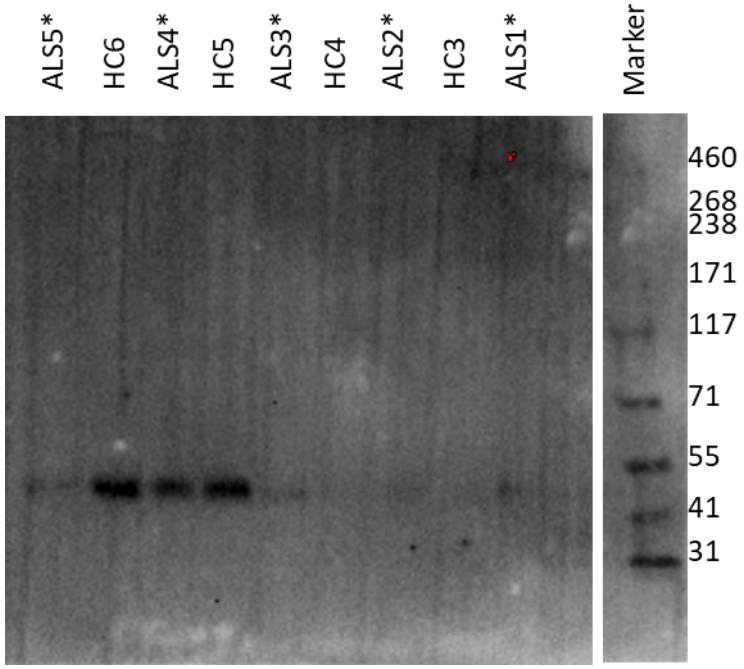


Figure 6.10. NfH profile in plasma NCHs digested with chymotrypsin.
Western blot against NfH following NCHs chymotrypsin digestion revealing a band at around 41 KDa for ALS4*, HC6 and HC5 samples only, indicating a positive but low degree of resistance to digestion with chymotrypsin for the NfH fraction included in NCHs.

As for trypsin digestion, after chymotrypsin digestion we have detected NfH bands only in three samples including ALS4*, HC6 and HC5 (Figure 6.10). This result is complementary to that of trypsin digestion, where aggregated NfH in ALS samples showed a higher degree of resistance than HC, which showed higher resistance to chymotrypsin (Figure 6.8 and Figure 6.10).

6.3.4 Calpain digestion

Calpain is a calcium dependent non-lysosomal cysteine protease ubiquitously expressed, with a wide range of targets and it is known to carry out an uncomplete hydrolysis of its targets because of the amino acid sequence, not well defined yet, and of the substrate's tertiary structure specificity (260–262). It has a wide range of substrates, including Nf, and it has been linked to axonal dysfunction and Nf degeneration (263). In addition, data produced by our research group include the determination of high level of calpain in CSF from ALS patients (unpublished data).

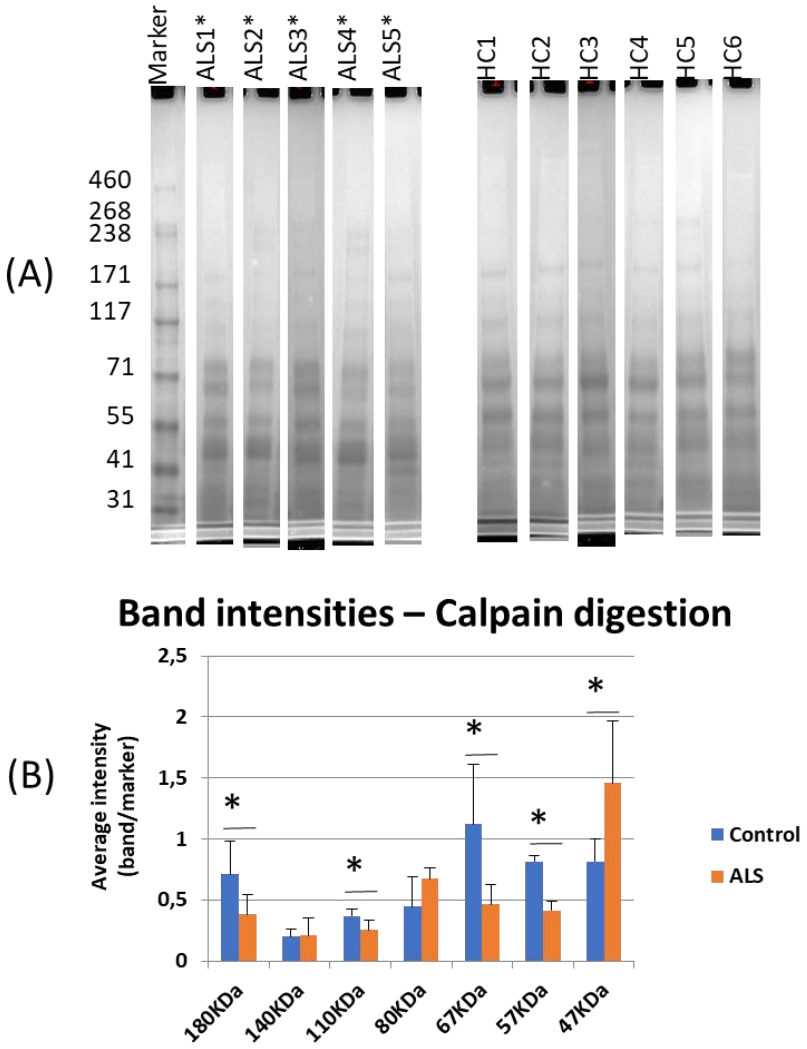


Figure 6.11. Plasma NCHs profile after calpain digestion.

(A) Plasma NCHs samples from ALS (ALS1*-5*) and HC (HC1-6) digested with calpain. (B) Quantitative analysis of the band profile. In this experiment, NCHs bands at 180 KDa were normalised with the 171 KDa Marker band, the 140 KDa bands with the 171 KDa Marker band,

the 110 KDa bands with the 117 KDa Marker band, the 80 KDa bands with the 71 KDa Marker band, the 67 KDa bands with the 71 KDa Marker band, the 57 KDa bands with the 55 KDa Marker bands and the 47 KDa bands the 55 KDa Marker band. Bands lower than 47KDa were not considered because it was difficult to discriminate from the background. *= p-Value < 0.05.

The digestion with calpain showed a very low specificity of the protease for plasma NCHs, with a significant degree of protease resistance (Figure 6.11A). In fact, all the samples showed after digestion several bands ranging from 180 KDa to 31 KDa, with ALS2*, ALS3*, ALS4*, HC4 and HC5 showing also bands at higher molecular weight (up to 238 KDa) (Figure 6.11A). Quantitative analysis of these bands between the ALS and HC groups showed that bands at 180, 110, 67 and 57 were more intense in HC NCHs compared to ALS (p= 0.036, 0.025, 0.021 and $2e^{-5}$ respectively), while the band at 47 KDa was more intense in ALS NCHs compared to HC (p= 0.045) (Figure 6.11B). Overall, for the bands represented in all samples after digestion, HC plasma NCHs seemed to be more resistant to calpain digestion than ALS plasma NCHs, showing a higher intensity for the bands detected (Figure 6.11B).

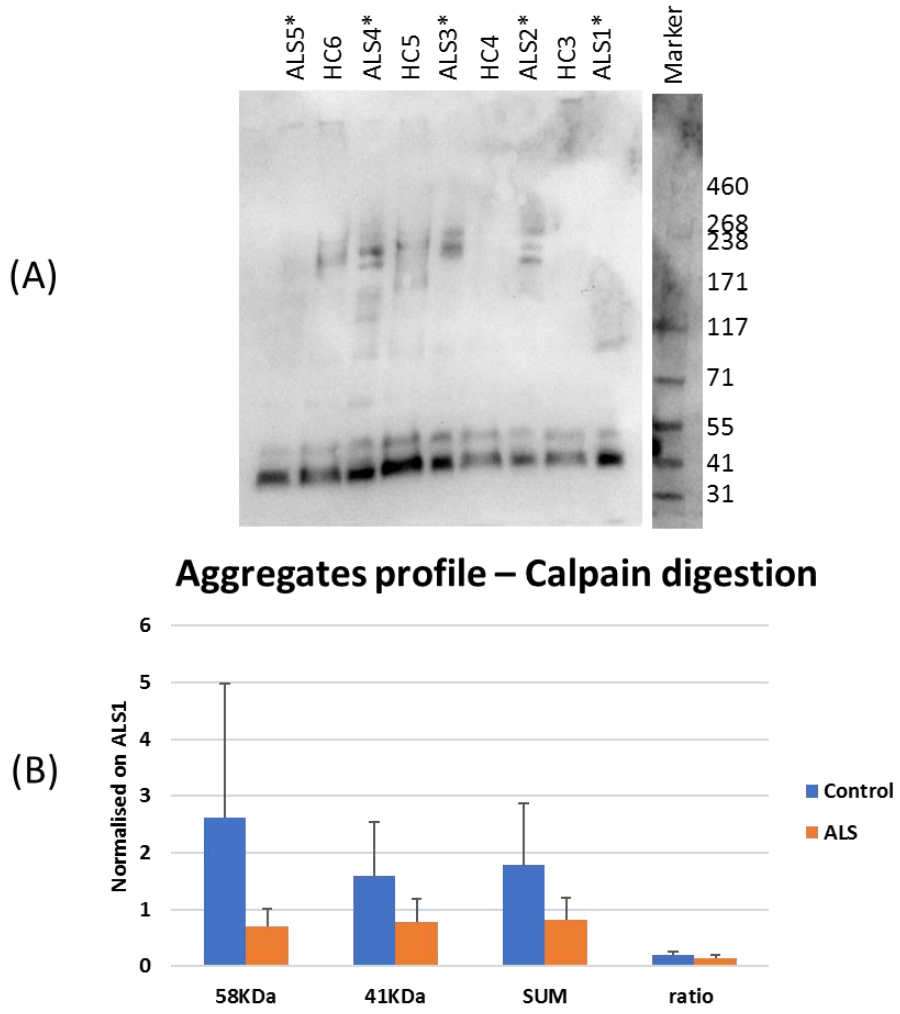


Figure 6.12. NfH profile in plasma NCHs digested with calpain.

(A) A) Western blot against NfH revealed two bands at 58KDa and 41KDa, shared across all the samples. HC6, ALS4*, HC5, ALS3* and ALS2* show other bands at higher MW, (B) Quantitative analysis of the bands detected including the sum of the 58 and 41 KDa bands (SUM) and the ratio between the 58 and 41 KDa bands (ratio) showed a higher, but not statistically significant, intensity in HC compared to ALS (p= 0.14). Band densities were normalised on ALS1* and corrected for the NfH content in each sample (Figure 6.6A).

NfH detection after digestion with calpain showed a wide range of immunoreactive peptides, mostly unique for each of the samples under investigation (Figure 6.12A). Only two peptides were shared across all the samples, one detected at 58 KDa and the other one at 41 KDa (Figure 6.12A). Band intensities after digestion depend on aggregated NfH susceptibility to digestion, but also on the NfH content of the sample. For this reason, to improve accuracy of the quantitative analysis on NfH digestion, the 58 KDa band intensities were corrected for NfH intensity detected at 460 and 268 KDa in the

original undigested sample, while the 41 KDa band intensities were corrected for the overall amount of undigested NfH (Figure 6.6A). With this correction, NfH density for 58 and 41 KDa bands as well as the sum of the two (SUM) was still higher in HC than ALS plasma NCHs (Figure 6.12B). However, these differences were not statistically significant (p= 0.18, 0.13 and 0.14 respectively). Also, ALS vs HC differences in the ratio between 58 and 41 KDa bands (ratio) showed a non-statistically significant higher value for HC compared to ALS (p= 0.11) (Figure 6.12B).

Despite a wide range of substrates, calpain is reported to have a very specific hydrolytic activity and the profile generated in plasma NCHs confirms this observation (Figure 6.11A). Also, the different composition of plasma NCHs across samples as reported in Chapter 5 would support our finding of an extreme variability in digestion products across samples. Moreover, the different digestion patterns seen in NfH immunodetection after calpain digestion (together with the well-known calpain specificity to different three-dimensional structures of target substrates) would suggest that NCHs and their NfH content enriched from ALS and HC have different conformations. This result puts more emphasis on the work done so far on imaging by transmission electron microscopy (TEM). The higher level of NfH reactive bands in HC plasma NCHs, compared to ALS, would suggest a lower level of phosphorylation of this protein in HC (33), considering that phosphorylation is a well-known factor affecting proteolysis.

6.3.5 Enterokinase digestion

Enterokinase (also called enteropeptidase) cleavage site specificity is aspartate-aspartate-aspartate-aspartate-lysine and it is known to convert trypsinogen to trypsin. It has been shown to be active on proenzymes like chymotrypsinogen, procarboxypeptidases and proelastases and to act specifically on NfH (38,264).

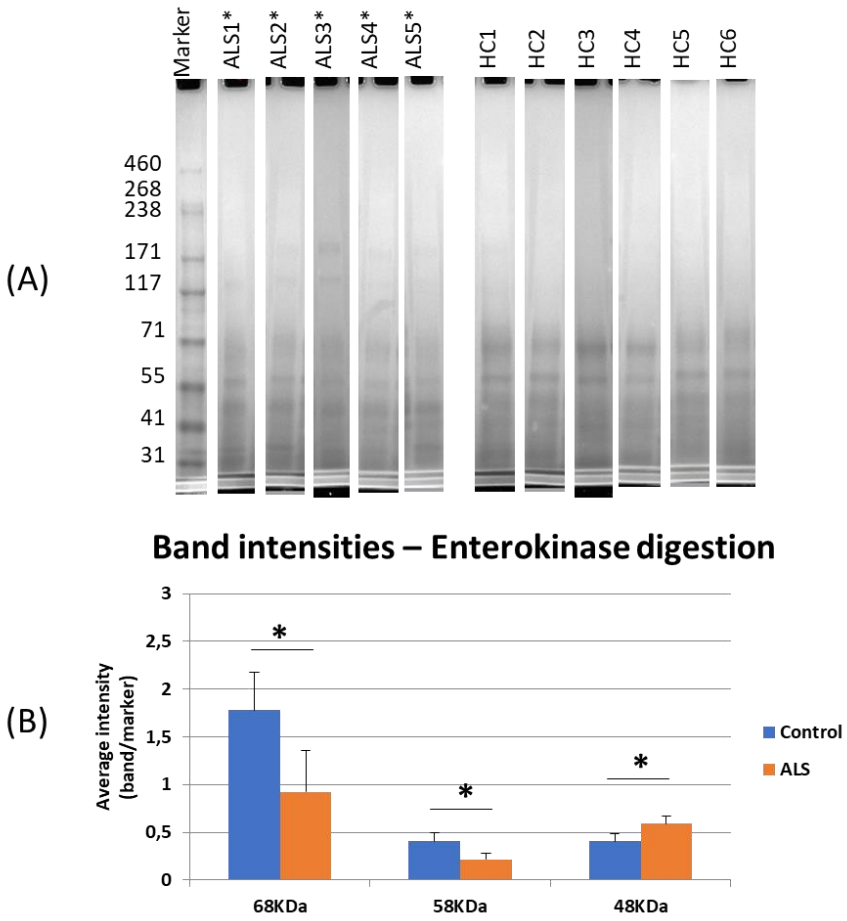


Figure 6.13. Plasma NCHs profile after enterokinase digestion.

(A) Plasma NCHs samples from ALS patients (ALS1*-5*) and HC (HC1-6) digested with enterokinase. (B) Quantitative analysis of the band profile. In this experiment, bands at 68 KDa were normalised with the 71 KDa Marker band, the 58 KDa bands with the 55 KDa Marker band, the 48 KDa bands with the 55 KDa Marker band. Bands below 48 KDa were not considered because it was difficult to discriminate them from the background. *= p-Value < 0.05.

Despite the known effect of enterokinase in the activation of trypsin and chymotrypsin, plasma NCHs digestion by enterokinase did not resemble the digestion profiles generated by these two enzymes in the same NCHs (Figure 6.7 and Figure 6.9). Enterokinase digestion of plasma NCHs generated three bands at 68, 58 and 48 KDa in all samples tested, with ALS2*, ALS3* and ALS4* showed also bands at 117 and 171 KDa (Figure 6.13A). Quantitative analysis of these bands between ALS and HC samples showed that bands at 68 and 58 were more intense in the HC NCHs compared to ALS (p= 0.009 and 0.004 respectively), while the band at 48 KDa was more intense in the ALS NCHs compared to HC (p= 0.005) (Figure 6.13B).

The NfH immunodetection profile after digestion revealed two different profiles between ALS and HC plasma NCHs samples (Figure 6.14A). All the samples shared a band at 50 KDa, but the ALS samples only also showed bands at 171 (not for ALS1*) and 31 KDa, not expressed by the HC samples (Figure 6.14A). The only band shared across all the plasma NCHs samples at 50 KDa was equally expressed between ALS and HC samples (Figure 6.14B).

Based on the data outlined in this paragraph, enterokinase seems to be very active in the test NCHs. In line with previous report of the proteolytic activity of enterokinase on NfH, the NfH profile in the ALS group suggests the presence of disease-specific changes in the NfH protein in NCHs, which may be at both biochemical (e.g. phosphorylation or glycosylation) or structural level. Interestingly, the peptide observed at 31 KDa in our experiment of NCHs eneterokinase digestion in ALS samples could be in line with the findings from Petzold et al., who have reported the presence of two non-phosphorylated NfH peptides (35–39 kDa, called $NfH_{835\text{-}1026}$, and 25–28 kDa, called $NfH_{852\text{-}986}$, fragments of similar size (38). The other immunoreactive bands identified in this study, including the one at 171 KDa specific for ALS seem not to have been previously reported in the literature.

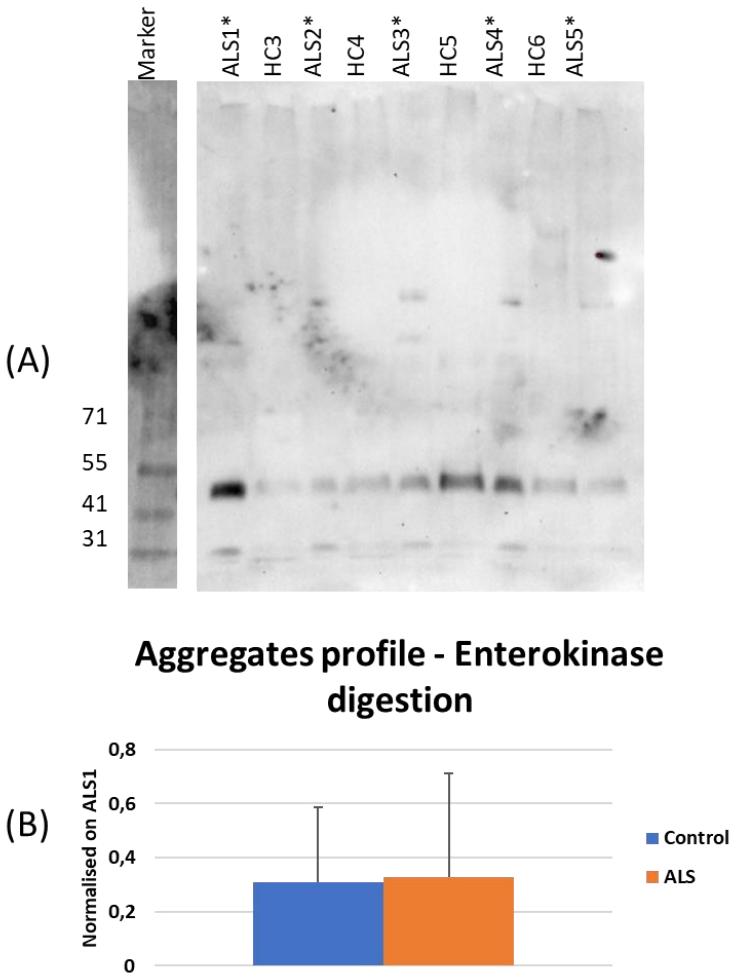


Figure 6.14. NfH profile in plasma NCHs digested with enterokinase.

(A) Western blot against NfH revealed one band at 50 KDa shared across all the samples. With the exclusion of ALS1*, all ALS samples also show two more bands at 171KDa and 31KDa not detected in HC samples. (B) Quantitative analysis of the band profile. Band densities were normalised on ALS1* and corrected for the NfH content in each sample (Figure 6.6A).

6.3.6 ALS brain NCHs digestion patterns

The resistance to protease digestion was tested also on brain NCHs, using chymotrypsin, calpain and enterokinase (Figure 6.15). Trypsin digestion on this sample was already tested in Paragraph 4.6. The gel profiles (Figure 6.15A) after protease treatment showed a complete digestion with chymotrypsin (lane 2), while digestion with enterokinase and calpain (lane 3 and 4 respectively) showed a darker area between 55 and 41 KDa, suggesting accumulation of the hydrolytic peptides.

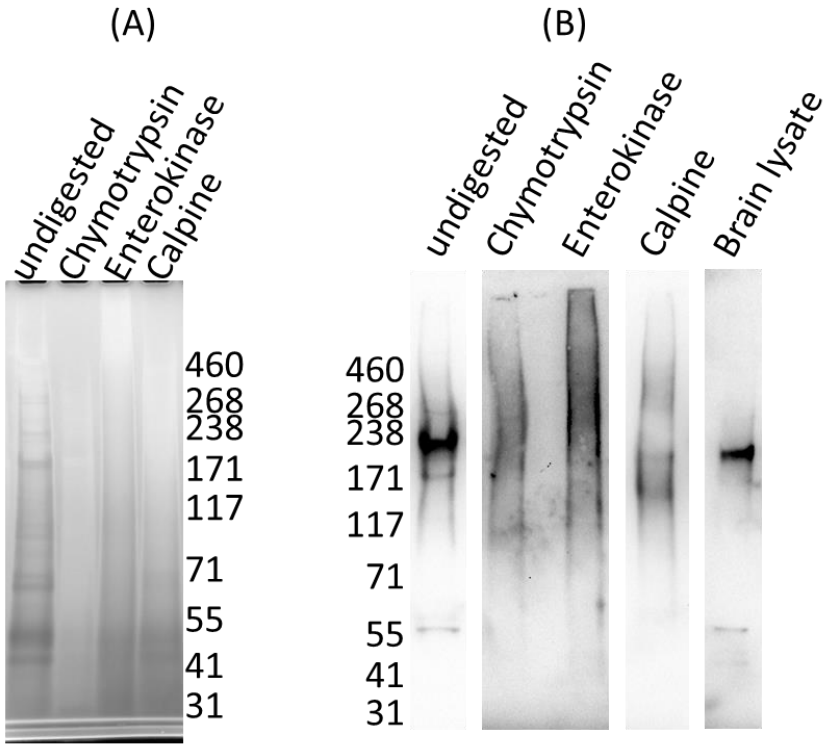


Figure 6.15. ALS brain NCHs digestion with proteases.

Zinc staining (A) and anti-NfH western blot (B) of brain NCHs digested with chymotrypsin, enterokinase and calpain and of the same undigested samples. In (B), pictures of the same gel were taken at different exposure times to maximise band visualization (lane 1 at 10.1 seconds, lane 4 and 5 at 58.4 seconds and lane 2 and 3 at 278.8 seconds).

NfH in brain NCHs showed very faint bands after digestion with the three enzymes suggesting a low resistance degree to digestion for this protein in the aggregated forms in brain (Figure 6.15B). Interestingly, chymotrypsin and enterokinase digestions (lane 2 and 3 respectively) generated two immunoreactive areas, rather than distinct bands, at higher molecular weight than the NfH detected in undigested brain lysate and NCHs (lane 5 and 1 respectively, Figure 6.15B). Calpain digestion showed two faint bands at about 171 KDa. It was the enzyme with the lowest hydrolytic activity on aggregates and NfH among the three tested in this experiment (lane 4, Figure 6.15B).

The presence of NfH species in digested brain NCHs at higher MW than that exhibited by undigested brain NCHs is difficult to interpret. This may be related to the dynamic of aggregation of NfH and of other co-aggregating proteins at high molecular weight and to their behaviour in the presence of proteases (Figure

6.15B). Alternatively, this could be due to the general low solubility of the aggregates in non-urea buffers (see Paragraph 2.6).

6.4 Conclusion

In this Chapter, we have built on our initial data to provide a comprehensive characterization of plasma and brain NCHs using imaging by transmission electron microscopy (TEM) and testing proteases resistance.

A relevant finding in our TEM work to characterize the morphology of plasma and brain aggregates is the identification of small filament-like particles similar to those detected in brain NCHs in the NCHs enriched plasma fraction (Figure 4.10 and Figure 6.3). This may suggest a cross talk between brain and systemic circulation as these macromolecular structures may find their way from the CNS to the peripheral circulation. However, this observation may also indicate that proteins like Nf, perhaps in a disordered state, may have a similar seeding effect in brain and plasma although the composition of co-aggregated proteins may be different in different matrices. Against the hypothesis that brain and plasma NCHs are closely related and perhaps also biologically similar in the formation of these aggregates is the observation that the digestion patterns of plasma and brain NCHs were completely different when using a number of hydrolytic enzymes (Figure 6.9-Figure 6.15). This discrepancy may point towards an over-interpretation of the TEM data, where morphology may not necessarily have the biological and functional meaning, particularly when it is not possible to speculate any further on the microstructure of these formations. However, it is also important to recognise that our TMT calibrator results show how a substantial number of proteins which are part of the neuronal and axonal network are also detected within peripheral aggregates, a finding that suggest a strong biological link between brain and circulating macromolecules. To further investigate these particles from a morphological point of view, we will have to rely on NCHs enrichment on a larger number of plasma and brain samples and consider inclusion of positive controls like for example prions particles.

Despite the presence of commonalities in the overall protein composition, the study of the protease digestion patterns in brain and plasma NCHs (Figure 6.15A) revealed a completely different activity of the hydrolytic enzymes used for the digestion experiments. While brain NCHs digestion with chymotrypsin,

enterokinase and calpain did not show any relevant form of resistance to digestion, plasma NCHs treated with the same proteases, and in particular with calpain and enterokinase (Figure 6.11 and Figure 6.13), generated a more complex pattern of digestion supporting the idea that the plasma protein mix including those of potential brain origin, acquire different biochemical properties once in circulation. Our data may also support the development of a new concept of biomarker where circulating NCHs may be a disease signature not only in virtue of their structure/composition but also for their functional role. Experiments on cell viability should be undertaken to evaluate possible toxicity of these aggregates.

Plasma NCHs digestion experiments showed a substantially different interaction between aggregates and proteases in ALS and healthy state. The digestion profiles showed both quantitative (Figure 6.11B and Figure 6.13B; i.e. the same bands in the gel digestion profiles showed different intensities across samples) and qualitative differences, where condition-specific digestion patterns showed bands appearing in one experimental group only (Figure 6.13A). The same condition-specific digestion profiles were identified when testing NfH susceptibility to protease digestion. NfH showed a degree of protease resistance in some of the samples particularly with calpain and enterokinase digestion and clear differences were detected between the ALS and the HC samples. While no definitive conclusion can be drawn from the analysis of relatively small number of samples, it is tempting to speculate that these features may be linked to the pathogenesis of the disease or play an important role in its development. Extending this investigation to a large number of samples to include, for example, non-neurological controls with other human conditions may be a way forward in the characterization of this finding. In addition, since resistance to protease digestion of NCHs enriched from different tissues may be the consequence of different PTM loads and types, which are well-known to condition structure and biochemical properties of molecules, it will be important to undertake proteomic studies which are geared towards the detection of the spectrum of relevant PTMs.

7 Summary and conclusion

In this thesis, I have developed a multi-modal experimental approach to study protein aggregation in blood in relation to the development of neurodegeneration. I have focused on this biological phenomenon after observing the behavior of axonal proteins like neurofilaments when they reach the blood stream after their release from degenerating neurons and axons. We have therefore concentrated on Neurofilament-Containing Heteroaggregates (NCHs) in blood and compared these formations to the well-known process of protein aggregation occurring in brain. What we find is that circulating NCHs retain molecular features of brain aggregation and they acquire biochemical properties that may be specific of pathological states characterized by neurodegeneration. Our data strongly support further investigations into these biological substrates as they may be relevant in the development of novel disease biomarkers and therapeutics for neurodegenerative diseases.

I have first tested methodologies for the enrichment of NCHs from blood and brain tissue (Chapter 3). Then NCHs composition was studied by mass spectrometry (MS)-based proteomics. NCHs composition was characterized in blood and brain samples from individuals with amyotrophic lateral sclerosis (ALS), a fatal neurodegenerative disorder, and in plasma samples from matched healthy controls (HC). Functional analysis of the NCHs proteomes from healthy and disease states has highlighted the presence of biochemical differences in protein degradation (proteasome activity) and energy metabolism (glycolysis/gluconeogenesis, carbon metabolism and pentose phosphate) already described in ALS literature (Chapter 4).

After having confirmed that NCHs are able to depict the biochemical and pathophysiological changes that may accompany the development of a neurodegenerative disorder, a quantitative analysis with the innovative TMTcalibrator™ workflow was undertaken using plasma NCHs enriched from six ALS patients and six matched HC (Chapter 5). This platform has the unique added value of being able to analyse tissues and fluids in the same round of experiments, dissecting fluid proteins that are also expressed in tissue. Thanks to this novel experimental approach, we have been able to detected and quantify

almost 5000 brain-derived proteins in plasma NCHs, with 285 of these showing a significant level of regulation ($p < 0.05$) in ALS compared to controls. Functional analysis has identified regulated pathways known to be involved in ALS pathology and in protein aggregates homeostasis including lysosome as well as lipoprotein and glycosaminoglycan metabolisms. Using immunodetection we have not been able to fully replicate the level of regulation in selected protein candidates obtained using the TMTcalibrator™ workflow. Taking into account the differences across different methodologies of protein detection, it will be certainly important to extend validation to a larger number of protein candidates that have emerged from our proteomic experiments.

Our hypothesis of NCHs representing a sound source of biomarkers in pathological condition is based not only on the MS-based proteomics data, but also on experiments in which we have tested their resistance to protease digestion (Chapter 6). The different digestion patterns in NCHs from ALS and HC observed using several proteases strengthen the hypothesis that these formations may go through disease-specificity biochemical changes, representing proteome differences in specific conditions, that could be ultimately exploited for the development of innovative disease biomarkers.

Ultimately, we believe that Neurofilament Heavy (NfH) may acquire a pro-aggregating tendence (Chapter 3) and as for many other proteins, this biological feature may be enhanced in neurodegenerative conditions. NfH is currently used as one of the possible biomarkers of ALS that is tested by simple plasma quantification. In this thesis, we have shown that NfH within heterogeneous aggregates has the potential to be be as informative as its soluble version. The relevance of different NfH species in their aggregated form was also highlighted by the experiments carried out with protease enzymes. In particular, enterokinase showed digestion-products specific for ALS plasma NCHs that were not present in HC plasma NCHs. The lack of peptides detected with a normal mass spectrometry analysis, probably due to the very low amount of these proteins in blood in respect of other more abundant proteins, indicates that immunoassays are still the best method for neurofilaments detection and quantification. However, the data showed in this thesis suggest the presence of specific isoforms of NfH that could be relevant for diagnosis, supporting the idea of specific studies

to evaluate the isoforms differences. This requires the development of specific methodologies for enrichment that would increase the chance of detection with other analytical instruments, such as mass spectrometry.

Another important data set generated in this project comes from the transmission electron microscopy (TEM) analysis of the NCHs fractions. These data somehow confirmed the validity of our NCHs enrichment protocol as a sound method for protein aggregates analysis (Chapter 3), showing macromolecular (mostly amorphous) formations in the extraction products. Despite in ALS literature there is no mention of prion-like structures in brain aggregates, TEM has also identified formations which may have resemblance to prion-like proteins (Chapter 4) (see figures in (165–167)). However, because of the absence of quality controls included in this set of experiments and the lack of references in literature, we cannot speculate any further and we will need further analysis and validation with a larger set of experiments.

The low number of samples tested in these experiments is not sufficient for a full evaluation of the diagnostic potential and of the involvement in the pathogenesis of ALS of NCHs. However, our work represents the ground for future validation of the biomarker candidates proposed in this thesis and the use of NCHs as surrogate of standard soluble molecules. In addition, testing these particles using established cell bioassays will give us information on the biological properties of these formation and eventual toxicity. This will be an important step towards understanding whether and how NCHs could be exploited for therapeutic purpose and to establish novel biomarkers.

The laborious methodology, the high volume of plasma required as well as the low high-throughput capacity of this approach does not suggest a possible application for NCHs in clinical practice. However, this project showed the potential of these particles for research studies aimed at biomarkers discovery, with the possibility to determine possible candidates that can be further developed for possible diagnostic tools in clinical practice.

In conclusion, combining a pioneering TMTcalibrator™ workflow with a range of biochemical and imaging techniques has been a valuable effort in the characterisation of a poorly understood phenomenon and in the discovery of a

potentially new field of investigation in the ever-growing medical and social emergency caused by neurodegenerative disorders.

8 References

1. Geisler N, Kaufmann E, Fischer S, Plessmann U, Weber K. Neurofilament architecture combines structural principles of intermediate filaments with carboxy-terminal extensions increasing in size between triplet proteins. Embo j. 1983/01/01. 1983;2(8):1295–302.

2. Angelides KJ, Smith KE, Takeda M. Assembly and exchange of intermediate filament proteins of neurons: neurofilaments are dynamic structures. J Cell Biol. 1989/04/01. 1989;108(4):1495–506.

3. Heins S, Wong PC, Muller S, Goldie K, Cleveland DW, Aebi U. The rod domain of NF-L determines neurofilament architecture, whereas the end domains specify filament assembly and network formation. J Cell Biol. 1993/12/01. 1993;123(6 Pt 1):1517–33.

4. Herrmann H, Haner M, Brettel M, Ku NO, Aebi U. Characterization of distinct early assembly units of different intermediate filament proteins. J Mol Biol. 1999/03/05. 1999;286(5):1403–20.

5. Szaro BG, Strong MJ. Post-transcriptional control of neurofilaments: New roles in development, regeneration and neurodegenerative disease. Trends Neurosci [Internet]. 2010;33(1):27–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19906448

6. Julien JP. Neurofilament functions in health and disease. Curr Opin Neurobiol. 1999/10/06. 1999;9(5):554–60.

7. Wagner OI, Ascano J, Tokito M, Leterrier JF, Janmey PA, Holzbaur EL. The interaction of neurofilaments with the microtubule motor cytoplasmic dynein. Mol Biol Cell. 2004/09/03. 2004;15(11):5092–100.

8. Mehta AD, Rock RS, Rief M, Spudich JA, Mooseker MS, Cheney RE. Myosin-V is a processive actin-based motor. Nature. 1999/08/17. 1999;400(6744):590–3.

9. Bocquet A, Berges R, Frank R, Robert P, Peterson AC, Eyer J. Neurofilaments bind tubulin and modulate its polymerization. J Neurosci. 2009/09/04. 2009;29(35):11043–54.

10. Goldenring JR, Lasher RS, Vallano ML, Ueda T, Naito S, Sternberger NH, et al. Association of synapsin I with neuronal cytoskeleton. Identification in cytoskeletal preparations in vitro and immunocytochemical localization in brain of synapsin I. J Biol Chem. 1986/06/25. 1986;261(18):8495–504.

11. Hirokawa N. Cross-linker system between neurofilaments, microtubules, and membranous organelles in frog axons revealed by the quick-freeze, deep-etching method. J Cell Biol. 1982/07/01. 1982;94(1):129–42.

12. Metuzals J, Mushynski WE. Electron microscope and experimental investigations of the neurofilamentous network in Deiters' neurons. Relationship with the cell surface and nuclear pores. J Cell Biol. 1974/06/01. 1974;61(3):701–22.

13. Friede RL, Samorajski T. Axon caliber related to neurofilaments and microtubules in sciatic nerve fibers of rats and mice. Anat Rec. 1970/08/01. 1970;167(4):379–87.

14. Zhu Q, Couillard-Despres S, Julien JP. Delayed maturation of regenerating myelinated axons in mice lacking neurofilaments. Exp Neurol. 1997/12/17. 1997;148(1):299–316.

15. Carden MJ, Trojanowski JQ, Schlaepfer WW, Lee VM. Two-stage expression of neurofilament polypeptides during rat neurogenesis with early establishment of adult phosphorylation patterns. J Neurosci [Internet]. 1987;7(11):3489–504. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3119790

16. Cuenca N, Fernandez E, de Juan J, Carreres J, Iniguez C. Postnatal development of microtubules and neurofilaments in the rat optic nerve: a quantitative study. J Comp Neurol. 1987/09/22. 1987;263(4):613–7.

17. Breen KC, Anderton BH. Temporal expression of neurofilament polypeptides in differentiating neuroblastoma cells. Neuroreport. 1991/01/01. 1991;2(1):21–4.

18. Xu Z, Marszalek JR, Lee MK, Wong PC, Folmer J, Crawford TO, et al. Subunit composition of neurofilaments specifies axonal diameter. J Cell Biol. 1996/06/01. 1996;133(5):1061–9.

19. Meier J, Couillard-Despres S, Jacomy H, Gravel C, Julien JP. Extra neurofilament NF-L subunits rescue motor neuron disease caused by overexpression of the human NF-H gene in mice. J Neuropathol Exp Neurol. 1999/10/09. 1999;58(10):1099–110.

20. Shen H, Barry DM, Garcia ML. Distal to proximal development of peripheral nerves requires the expression of neurofilament heavy. Neuroscience. 2010/07/17. 2010;170(1):16–21.

21. Grant P, Pant HC. Neurofilament protein synthesis and phosphorylation. J Neurocytol [Internet]. 2000;29(11–12):843–72. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11466475

22. Sihag RK, Inagaki M, Yamaguchi T, Shea TB, Pant HC. Role of phosphorylation on the structural dynamics and function of types III and IV intermediate filaments. Exp Cell Res. 2007/05/15. 2007;313(10):2098–109.

23. Nixon RA, Lewis SE. Differential turnover of phosphate groups on neurofilament subunits in mammalian neurons in vivo. J Biol Chem. 1986/12/15. 1986;261(35):16298–301.

24. Nixon RA, Lewis SE, Dahl D, Marotta CA, Drager UC. Early posttranslational modifications of the three neurofilament subunits in mouse retinal ganglion cells: neuronal sites and time course in relation to subunit polymerization and axonal transport. Brain Res Mol Brain Res. 1989/03/01. 1989;5(2):93–108.

25. Nixon RA, Lewis SE, Marotta CA. Posttranslational modification of

neurofilament proteins by phosphate during axoplasmic transport in retinal ganglion cell neurons. J Neurosci. 1987/04/01. 1987;7(4):1145–58.

26. Nixon RA, Paskevich PA, Sihag RK, Thayer CY. Phosphorylation on carboxyl terminus domains of neurofilament proteins in retinal ganglion cell neurons in vivo: influences on regional neurofilament accumulation, interneurofilament spacing, and axon caliber. J Cell Biol. 1994/08/01. 1994;126(4):1031–46.

27. Oblinger MM, Brady ST, McQuarrie IG, Lasek RJ. Cytotypic differences in the protein composition of the axonally transported cytoskeleton in mammalian neurons. J Neurosci. 1987/02/01. 1987;7(2):453–62.

28. Ackerley S, Thornhill P, Grierson AJ, Brownlees J, Anderton BH, Leigh PN, et al. Neurofilament heavy chain side arm phosphorylation regulates axonal transport of neurofilaments. J Cell Biol. 2003/05/14. 2003;161(3):489–95.

29. Perrot R, Eyer J. Neurofilaments: Properties, Functions, and Regulation. In: Dermietzel R, editor. The Cytoskeleton: Imaging, Isolation, and Interaction [Internet]. Totowa, NJ: Humana Press; 2013. p. 171–236. Available from: https://doi.org/10.1007/978-1-62703-266-7_9

30. Slawson C, Hart GW. Dynamic interplay between O-GlcNAc and O-phosphate: the sweet side of protein regulation. Curr opin struct biol. 2003/10/22. 2003;13(5):631–6.

31. Gou JP, Leterrier JF. Possible involvement of ubiquitination in neurofilament degradation. Biochem Biophys Res Commun. 1995/12/14. 1995;217(2):529–38.

32. Schlaepfer WW, Lee C, Lee VM, Zimmerman UJ. An immunoblot study of neurofilament degradation in situ and during calcium-activated proteolysis. J Neurochem. 1985/02/01. 1985;44(2):502–9.

33. Pant HC. Dephosphorylation of neurofilament proteins enhances their susceptibility to degradation by calpain. Biochem J [Internet]. 1988;256(2):665–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2851997

34. Wang Q, Song F, Zhang C, Zhao X, Zhu Z, Yu S, et al. Carboxyl-terminus of Hsc70 interacting protein mediates 2,5-hexanedione-induced neurofilament medium chain degradation. Biochem Pharmacol. 2011;

35. Balastik M, Ferraguti F, Pires-da Silva A, Lee TH, Alvarez-Bolado G, Lu KP, et al. Deficiency in ubiquitin ligase TRIM2 causes accumulation of neurofilament light chain and neurodegeneration. Proc Natl Acad Sci. 2008;

36. Chen JX, Sun YJ, Wang P, Long DX, Li W, Li L, et al. Induction of autophagy by TOCP in differentiated human neuroblastoma cells lead to degradation of cytoskeletal components and inhibition of neurite outgrowth. Toxicology. 2013;

37. Song F, Zhang Q, Kou R, Zou C, Gao Y, Xie K. 2,5-Hexanedione Altered the Degradation of Low-Molecular-Weight Neurofilament in Rat Nerve

Tissues. Food Chem Toxicol. 2012;

38. Petzold A, Tisdall MM, Girbes AR, Martinian L, Thom M, Kitchen N, et al. In vivo monitoring of neuronal loss in traumatic brain injury: A microdialysis study. Brain. 2011;

39. Scott D, Smith KE, O'Brien BJ, Angelides KJ. Characterization of mammalian neurofilament triplet proteins: Subunit stoichiometry and morphology of native and reconstituted filaments. J Biol Chem. 1985;

40. Zucchi E, Lu CH, Cho Y, Chang R, Adiutori R, Zubiri I, et al. A motor neuron strategy to save time and energy in neurodegeneration: adaptive protein stoichiometry. J Neurochem. 2018;

41. Noble W, Burns MP. Challenges in neurodegeneration research. Front Psychiatry. 2010;

42. Rachakonda V, Pan TH, Le WD. Biomarkers of neurodegenerative disorders: how good are they? Cell Res. 2004/11/13. 2004;14(5):347–58.

43. Bradley E. Incorporating biomarkers into clinical trial designs: points to consider. Nat Biotechnol. 2012/07/12. 2012;30(7):596–9.

44. Kim D, Kim YS, Shin DW, Park CS, Kang JH. Harnessing cerebrospinal fluid biomarkers in clinical trials for treating Alzheimer's and Parkinson's diseases: Potential and challenges. Journal of Clinical Neurology (Korea). 2016.

45. Hoglund K, Salter H. Molecular biomarkers of neurodegeneration. Expert Rev Mol Diagn. 2013/10/25. 2013;13(8):845–61.

46. Lu CH, Petzold A, Topping J, Allen K, Macdonald-Wallis C, Clarke J, et al. Plasma neurofilament heavy chain levels and disease progression in amyotrophic lateral sclerosis: insights from a longitudinal study. J Neurol Neurosurg Psychiatry [Internet]. 2015;86(5):565–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25009280

47. Gaiottino J, Norgren N, Dobson R, Topping J, Nissim A, Malaspina A, et al. Increased neurofilament light chain blood levels in neurodegenerative neurological diseases. PLoS One [Internet]. 2013;8(9):e75091. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24073237

48. Lu CH, Macdonald-Wallis C, Gray E, Pearce N, Petzold A, Norgren N, et al. Neurofilament light chain: A prognostic biomarker in amyotrophic lateral sclerosis. Neurology. 2015/05/03. 2015;84(22):2247–57.

49. Robelin L, Gonzalez De Aguilar JL. Blood biomarkers for amyotrophic lateral sclerosis: myth or reality? Biomed Res Int [Internet]. 2014;2014:525097. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24991560

50. Batada N, Reguly T, Breitkreutz A, Boucher L, Breitkreutz B, Hurst L, et al. Still stratus not altocumulus: further evidence against the date/party hub distinction. Plos Biol [Internet]. 2007;5. Available from: http://dx.doi.org/10.1371/journal.pbio.0050154

51. Garbuzova-Davis S, Hernandez-Ontiveros DG, Rodrigues MC, Haller E, Frisina-Deyo A, Mirtyl S, et al. Impaired blood-brain/spinal cord barrier in ALS patients. Brain Res [Internet]. 2012;1469:114–28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22750125

52. Absinta M, Ha SK, Nair G, Sati P, Luciano NJ, Palisoc M, et al. Human and nonhuman primate meninges harbor lymphatic vessels that can be visualized noninvasively by MRI. Elife. 2017;

53. Lu CH, Kalmar B, Malaspina A, Greensmith L, Petzold A. A method to solubilise protein aggregates for immunoassay quantification which overcomes the neurofilament “hook” effect. J Neurosci Methods [Internet]. 2011;195(2):143–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21134399

54. Bahmanyar S, Moreau-Dubois MC, Brown P, Cathala F, Gajdusek DC. Serum antibodies to neurofilament antigens in patients with neurological and other diseases and in healthy controls. J Neuroimmunol. 1983/10/01. 1983;5(2):191–6.

55. Ross CA, Poirier MA. What is the role of protein aggregation in neurodegeneration? Nat Rev Mol Cell Biol [Internet]. 2005;6(11):891–8. Available from: http://dx.doi.org/10.1038/nrm1742

56. Yang H, Hu HY. Sequestration of cellular interacting partners by protein aggregates: implication in a loss-of-function pathology. FEBS J [Internet]. 2016;283(20):3705–17. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27016044

57. Brangwynne CP. Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. Soft Matter [Internet]. 2011;7(7):3052–9. Available from: http://dx.doi.org/10.1039/C0SM00981D

58. David DC, Ollikainen N, Trinidad JC, Cary MP, Burlingame AL, Kenyon C. Widespread protein aggregation as an inherent part of aging in C. elegans. Plos Biol. 2010/08/17. 2010;8(8):e1000450.

59. Kaganovich D, Kopito R, Frydman J. Misfolded proteins partition between two distinct quality control compartments. Nature. 2008/08/30. 2008;454(7208):1088–95.

60. Wallace EW, Kear-Scott JL, Pilipenko E V, Schwartz MH, Laskowski PR, Rojek AE, et al. Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. Cell. 2015/09/12. 2015;162(6):1286–98.

61. Walther DM, Kasturi P, Zheng M, Pinkert S, Vecchi G, Ciryam P, et al. Widespread Proteome Remodeling and Aggregation in Aging C. elegans. Cell. 2015/05/11. 2015;161(4):919–32.

62. Rubinsztein DC. The roles of intracellular protein-degradation pathways in neurodegeneration. Nature [Internet]. 2006;443(7113):780–6. Available from: https://www.ncbi.nlm.nih.gov/pubmed/17051204

63. Buchberger A, Bukau B, Sommer T. Protein quality control in the cytosol

and the endoplasmic reticulum: brothers in arms. Mol Cell [Internet]. 2010;40(2):238–52. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20965419

64. Lim J, Yue Z. Neuronal aggregates: formation, clearance, and spreading. Dev Cell. 2015/02/25. 2015;32(4):491–501.

65. Li W, Prabakaran P, Chen W, Zhu Z, Feng Y, Dimitrov D. Antibody Aggregation: Insights from Sequence and Structure. Antibodies [Internet]. 2016;5(3):19. Available from: http://www.mdpi.com/2073-4468/5/3/19

66. Sahin E, Weiss WF th, Kroetsch AM, King KR, Kessler RK, Das TK, et al. Aggregation and pH-temperature phase behavior for aggregates of an IgG2 antibody. J Pharm Sci. 2012/01/17. 2012;101(5):1678–87.

67. Kolhe P, Amend E, Singh SK. Impact of freezing on pH of buffered solutions and consequences for monoclonal antibody aggregation. Biotechnol Prog. 2009/12/30. 2010;26(3):727–33.

68. Finn TE, Nunez AC, Sunde M, Easterbrook-Smith SB. Serum albumin prevents protein aggregation and amyloid formation and retains chaperone-like activity in the presence of physiological ligands. J Biol Chem. 2012/05/03. 2012;287(25):21530–40.

69. Xia K, Trasatti H, Wymer JP, Colon W. Increased levels of hyper-stable protein aggregates in plasma of older adults. Age [Internet]. 2016;38(3):56. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27179971

70. Corbo M, Hays AP. Peripherin and neurofilament protein coexist in spinal spheroids of motor neuron disease. J Neuropathol Exp Neurol. 1992/09/01. 1992;51(5):531–7.

71. Vogel P, Gabriel M, Goebel HH, Dyck PJ. Hereditary motor sensory neuropathy type II with neurofilament accumulation: new finding or new disorder? Ann Neurol. 1985/05/01. 1985;17(5):455–61.

72. Galloway PG, Mulvihill P, Perry G. Filaments of Lewy bodies contain insoluble cytoskeletal elements. Am J Pathol. 1992/04/01. 1992;140(4):809–22.

73. Bendotti C, Atzori C, Piva R, Tortarolo M, Strong MJ, DeBiasi S, et al. Activated p38MAPK is a novel component of the intracellular inclusions found in human amyotrophic lateral sclerosis and mutant SOD1 transgenic mice. J Neuropathol Exp Neurol [Internet]. 2004;63(2):113–9. Available from: https://www.ncbi.nlm.nih.gov/pubmed/14989597

74. Hirsch L, Jette N, Frolkis A, Steeves T, Pringsheim T. The Incidence of Parkinson's Disease: A Systematic Review and Meta-Analysis. Neuroepidemiology. 2016.

75. Association A, Alzheimer's Association, Association A. 2015 Alzheimer's disease facts and figures. Alzheimers Dement. 2015;

76. Arthur KC, Calvo A, Price TR, Geiger JT, Chiò A, Traynor BJ. Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat Commun.

2016;

77. Hardiman O, Al-Chalabi A, Brayne C, Beghi E, Van Den Berg LH, Chio A, et al. The changing picture of amyotrophic lateral sclerosis: Lessons from European registers. Journal of Neurology, Neurosurgery and Psychiatry. 2017.

78. Hardiman O, van den Berg LH, Kiernan MC. Clinical diagnosis and management of amyotrophic lateral sclerosis. Nat Rev Neurol [Internet]. 2011;7(11):639–49. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21989247

79. Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. Nat Rev Neurol. 2014/10/15. 2014;10(11):661–70.

80. Polymenidou M, Cleveland DW. The seeds of neurodegeneration: prion-like spreading in ALS. Cell [Internet]. 2011 Oct 28 [cited 2018 Aug 30];147(3):498–508. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22036560

81. Lee S, Kim HJ. Prion-like Mechanism in Amyotrophic Lateral Sclerosis: are Protein Aggregates the Key? Exp Neurobiol [Internet]. 2015;24(1):1–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25792864

82. Blokhuis AM, Groen EJ, Koppers M, van den Berg LH, Pasterkamp RJ. Protein aggregation in amyotrophic lateral sclerosis. Acta Neuropathol [Internet]. 2013;125(6):777–94. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23673820

83. Chou SM, Wang HS, Taniguchi A, Bucala R. Advanced glycation endproducts in neurofilament conglomeration of motoneurons in familial and sporadic amyotrophic lateral sclerosis. Mol Med. 1998/06/27. 1998;4(5):324–32.

84. Crow JP, Ye YZ, Strong M, Kirk M, Barnes S, Beckman JS. Superoxide dismutase catalyzes nitration of tyrosines by peroxynitrite in the rod and head domains of neurofilament-L. J Neurochem. 1998/02/12. 1997;69(5):1945–53.

85. Renton AE, Chio A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci [Internet]. 2014;17(1):17–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24369373

86. Peters OM, Ghasemi M, Brown Jr. RH. Emerging mechanisms of molecular pathology in ALS. J Clin Invest. 2015/05/02. 2015;125(5):1767–79.

87. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci. 1999/12/14. 1999;169(1–2):13–21.

88. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El

Escorial “Clinical limits of amyotrophic lateral sclerosis” workshop contributors. J Neurol Sci. 1994/07/01. 1994;124 Suppl:96–107.

89. Craft GE, Chen A, Nairn AC. Recent advances in quantitative neuroproteomics. Methods. 2013/04/30. 2013;61(3):186–218.

90. Stasyk T, Huber LA. Zooming in: fractionation strategies in proteomics. Proteomics. 2004/11/13. 2004;4(12):3704–16.

91. Amunugama R, Jones R, Ford M, Allen D. Bottom-Up Mass Spectrometry–Based Proteomics as an Investigative Analytical Tool for Discovery and Quantification of Proteins in Biological Samples. Adv Wound Care [Internet]. 2013;2(9):549–57. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3842888/

92. Nesvizhskii AI, Vitek O, Aebersold R. Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods. 2007/09/29. 2007;4(10):787–97.

93. Russell CL, Mitra V, Hansson K, Blennow K, Gobom J, Zetterberg H, et al. Comprehensive Quantitative Profiling of Tau and Phosphorylated Tau Peptides in Cerebrospinal Fluid by Mass Spectrometry Provides New Biomarker Candidates. Iqbal K, editor. J Alzheimer’s Dis [Internet]. 2016/09/17. 2016 Nov 1;55(1):303–13. Available from: http://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/JAD-160633

94. The ALS Association. Facts you should know [Internet]. June. 2016 [cited 2018 Aug 30]. Available from: http://www.alsa.org/about-als/facts-you-should-know.html

95. Armon C, Moses D. Linear estimates of rates of disease progression as predictors of survival in patients with ALS entering clinical trials. In: Journal of the Neurological Sciences. 1998.

96. Lane A, Stanley CJ, Dealler S, Wilson SM. Polymeric ligands with specificity for aggregated prion proteins. patent. 2003;PCT/GB03/00858.

97. Greenberg SG, Davies P. A preparation of Alzheimer paired helical filaments that displays distinct tau proteins by polyacrylamide gel electrophoresis. Proc Natl Acad Sci U S A. 1990;

98. Nesvizhskii AI, Aebersold R. Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. Mol Cell Proteomics. 2005;

99. Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Stat. 1992;

100. Polymenidou M, Cleveland DW. The seeds of neurodegeneration: prion-like spreading in ALS. Cell [Internet]. 2011;147(3):498–508. Available from: https://www.ncbi.nlm.nih.gov/pubmed/22036560

101. Lu C-H, Petzold A, Kalmar B, Dick J, Malaspina A, Greensmith L. Plasma Neurofilament Heavy Chain Levels Correlate to Markers of Late Stage Disease Progression and Treatment Response in SOD1(G93A) Mice that

Model ALS. PLoS One [Internet]. 2012;7(7):e40998. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3397981/

102. Berkowitz SA. Role of analytical ultracentrifugation in assessing the aggregation of protein biopharmaceuticals. AAPS J [Internet]. 2006;8(3):E590–605. Available from: http://link.springer.com/10.1208/aapsj080368

103. Uchiyama S, Noda M, Krayukhina E. Sedimentation velocity analytical ultracentrifugation for characterization of therapeutic antibodies. Biophys Rev. 2018;10(2):259–69.

104. Tebbenkamp AT, Borchelt DR. Protein aggregate characterization in models of neurodegenerative disease. Methods Mol Biol [Internet]. 2009;566:85–91. Available from: https://www.ncbi.nlm.nih.gov/pubmed/20058166

105. Hutornojs V, Niedre-Otomere B, Kozlovska T, Zajakina A. Comparison of ultracentrifugation methods for concentration of recombinant alphaviruses: sucrose and iodixanol cushions. Environ Exp Biol [Internet]. 2012;10:117–23. Available from: http://eeb.lu.lv/EEB/201212/EEB_10_Hutornojs.pdf

106. Kalra H, Adda CG, Liem M, Ang CS, Mechler A, Simpson RJ, et al. Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. Proteomics. 2013;

107. Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res [Internet]. 2003;31(1):334–41. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC165562/

108. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 2005/06/28. 2005;33(Web Server issue):W741-8.

109. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. In: Journal of the Royal Statistical Society. 1995.

110. Benn CL, Butler R, Mariner L, Nixon J, Moffitt H, Mielcarek M, et al. Genetic knock-down of HDAC7 does not ameliorate disease pathogenesis in the R6/2 mouse model of Huntington's disease. PLoS One. 2009/06/02. 2009;4(6):e5747.

111. Sathasivam K, Lane A, Legleiter J, Warley A, Woodman B, Finkbeiner S, et al. Identical oligomeric and fibrillar structures captured from the brains of R6/2 and knock-in mouse models of Huntington's disease. Hum Mol Genet. 2009/10/15. 2010;19(1):65–78.

112. Scherzinger E, Lurz R, Turmaine M, Mangiarini L, Hollenbach B, Hasenbank R, et al. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates in vitro and in vivo. Cell. 1997/08/08.

1997;90(3):549–58.

113. Wanker EE, Scherzinger E, Heiser V, Sittler A, Eickhoff H, Lehrach H. Membrane filter assay for detection of amyloid-like polyglutamine-containing protein aggregates. Methods Enzym. 1999/10/03. 1999;309:375–86.

114. Chang E, Kuret J. Detection and quantification of tau aggregation using a membrane filter assay. Anal Biochem. 2007/10/24. 2008;373(2):330–6.

115. Ayyadevara S, Balasubramaniam M, Parcon PA, Barger SW, Griffin WS, Alla R, et al. Proteins that mediate protein aggregation and cytotoxicity distinguish Alzheimer's hippocampus from normal controls. Aging Cell. 2016/07/28. 2016;15(5):924–39.

116. Dekker AD, Fortea J, Blesa R, De Deyn PP. Cerebrospinal fluid biomarkers for Alzheimer's disease in Down syndrome. Alzheimers Dement (Amst). 2017/04/18. 2017;8:1–10.

117. Arvinte T, Palais C, Green-Trexler E, Gregory S, Mach H, Narasimhan C, et al. Aggregation of biopharmaceuticals in human plasma and human serum: implications for drug research and development. MAbs. 2013/04/11. 2013;5(3):491–500.

118. Demeule B, Palais C, Machaidze G, Gurny R, Arvinte T. New methods allowing the detection of protein aggregates: a case study on trastuzumab. MAbs. 2010/01/12. 2009;1(2):142–50.

119. Betts JC, Blackstock WP, Ward MA, Anderton BH. Identification of phosphorylation sites on neurofilament proteins by nanoelectrospray mass spectrometry. J Biol Chem [Internet]. 1997;272(20):12922–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9148897

120. Dong DL, Xu ZS, Chevrier MR, Cotter RJ, Cleveland DW, Hart GW. Glycosylation of mammalian neurofilaments. Localization of multiple O-linked N-acetylglucosamine moieties on neurofilament polypeptides L and M. J Biol Chem [Internet]. 1993;268(22):16679–87. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8344946

121. Dale JM, Garcia ML. Neurofilament Phosphorylation during Development and Disease: Which Came First, the Phosphorylation or the Accumulation? J Amino Acids. 2012;

122. Dong DLY, Xu ZS, Hart GW, Cleveland DW. Cytoplasmic O-GlcNAc modification of the head domain and the KSP repeat motif of the neurofilament protein neurofilament-H. J Biol Chem. 1996;

123. Xu ZS, Liu WS, Willard MB. Identification of six phosphorylation sites in the COOH-terminal tail region of the rat neurofilament protein M. J Biol Chem. 1992;

124. Filipe V, Hawe A, Jiskoot W. Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. Pharm Res. 2010/03/06. 2010;27(5):796–810.

125. Ross CA, Poirier MA. Protein aggregation and neurodegenerative disease. Nat Med. 2004;

126. Adiutori R, Aarum J, Zubiri I, Bremang M, Jung S, Sheer D, et al. The proteome of neurofilament-containing protein aggregates in blood. Biochem Biophys Rep. 2018/06/07. 2018;14:168–77.

127. McCombe PA, Pfluger C, Singh P, Lim CYH, Airey C, Henderson RD. Serial measurements of phosphorylated neurofilament-heavy in the serum of subjects with amyotrophic lateral sclerosis. J Neurol Sci. 2015;

128. Boylan K, Yang C, Crook J, Overstreet K, Heckman M, Wang Y, et al. Immunoreactivity of the phosphorylated axonal neurofilament H subunit (pNF-H) in blood of ALS model rodents and ALS patients: Evaluation of blood pNF-H as a potential ALS biomarker. J Neurochem. 2009;

129. Duerkop M, Berger E, Dürauer A, Jungbauer A. Impact of Cavitation, High Shear Stress and Air/Liquid Interfaces on Protein Aggregation. Biotechnol J. 2018;

130. Kajak-Siemaszko K, Aubry L, Peyrin F, Bax ML, Gatellier P, Astruc T, et al. Characterization of protein aggregates following a heating and freezing process. Food Res Int. 2011;

131. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999/12/11. 2000;28(1):27–30.

132. Zhang B, Kirov S, Snoddy J. WebGestalt: An integrated system for exploring gene sets in various biological contexts. Nucleic Acids Res. 2005;33(SUPPL. 2):741–8.

133. Ibstedt S, Sideri TC, Grant CM, Tamás MJ. Global analysis of protein aggregation in yeast during physiological conditions and arsenite stress. Vol. 3, Biology Open. Bidder Building, 140 Cowley Road, Cambridge, CB4 0DL, UK; 2014. p. 913–23.

134. Bocharova N, Chave-Cox R, Sokolov S, Knorre D, Severin F. Protein aggregation and neurodegeneration: clues from a yeast model of Huntington's disease. Biochemistry (Mosc). 2009 Feb;74(2):231–4.

135. An L, Harrison PM. The evolutionary scope and neurological disease linkage of yeast-prion-like proteins in humans. Biol Direct. 2016;11(1).

136. Weids AJ, Ibstedt S, Tamás MJ, Grant CM. Distinct stress conditions result in aggregation of proteins with similar properties. Sci Rep [Internet]. 2016;6:24554. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27086931

137. Liu S, Hossinger A, Göbbels S, Vorberg IM. Prions on the run: How extracellular vesicles serve as delivery vehicles for self-templating protein aggregates. Prion [Internet]. 2017/04/14. 2017 Mar 4;11(2):98–112. Available from: https://www.tandfonline.com/doi/full/10.1080/19336896.2017.1306162

138. Carija A, Navarro S, de Groot NS, Ventura S. Protein aggregation into

insoluble deposits protects from oxidative stress. Redox Biol [Internet]. 2017;12:699–711. Available from: https://www.ncbi.nlm.nih.gov/pubmed/28410533

139. Harbi D, Kumar M, Harrison PM. LPS-annotate: complete annotation of compositionally biased regions in the protein knowledgebase. Database (Oxford). 2011;

140. Ross ED, Maclea KS, Anderson C, Ben-Hur A. A bioinformatics method for identifying Q/N-rich prion-like domains in proteins. Methods Mol Biol. 2013;1017:219–28.

141. Lancaster AK, Nutter-Upham A, Lindquist S, King OD. PLAAC: A web and command-line application to identify proteins with prion-like amino acid composition. Bioinformatics. 2014;

142. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;

143. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982 May;157(1):105–32.

144. Hock EM, Polymenidou M. Prion-like propagation as a pathogenic principle in frontotemporal dementia. J Neurochem [Internet]. 2016;138 Suppl 1:163–83. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27502124

145. Lee S, Kim H-J. Prion-like Mechanism in Amyotrophic Lateral Sclerosis: are Protein Aggregates the Key? Exp Neurobiol [Internet]. 2015 Mar [cited 2018 Aug 30];24(1):1–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25792864

146. McKinley MP, Bolton DC, Prusiner SB. A protease-resistant protein is a structural component of the scrapie prion. Cell. 1983/11/01. 1983;35(1):57–62.

147. Julien JP, Mushynski WE. The distribution of phosphorylation sites among identified proteolytic fragments of mammalian neurofilaments. J Biol Chem. 1983;

148. Yuan A, Rao M V., Veeranna, Nixon RA. Neurofilaments at a glance. J Cell Sci. 2012;

149. Kametani F, Obi T, Shishido T, Akatsu H, Murayama S, Saito Y, et al. Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral sclerosis brains. Sci Rep. 2016/03/17. 2016;6:23281.

150. Song C, Guo J, Liu Y, Tang B. Autophagy and Its Comprehensive Impact on ALS. Int J Neurosci [Internet]. 2012 Oct 26;122(12):695–703. Available from: http://www.tandfonline.com/doi/full/10.3109/00207454.2012.714430

151. Cipolat Mis MS, Brajkovic S, Frattini E, Di Fonzo A, Corti S. Autophagy in motor neuron disease: Key pathogenetic mechanisms and therapeutic targets. Mol Cell Neurosci [Internet]. 2016 Apr 1 [cited 2018 Oct 17];72:84–90. Available from:

https://www.sciencedirect.com/science/article/abs/pii/S1044743116300124

152. Wilkinson CRM, Seeger M, Hartmann-Petersen R, Stone M, Wallace M, Semple C, et al. Proteins containing the UBA domain are able to bind to multi-ubiquitin chains. Nat Cell Biol. 2001;

153. Raasi S, Orlov I, Fleming KG, Pickart CM. Binding of polyubiquitin chains to ubiquitin-associated (UBA) domains of HHR23A. J Mol Biol. 2004;

154. Johnston JA, Ward CL, Kopito RR. Aggresomes: A cellular response to misfolded proteins. J Cell Biol. 1998;

155. Lelouard H, Gatti E, Cappello F, Gresser O, Camosseto V, Pierre P. Transient aggregation of ubiquitinated proteins during dendritic cell maturation. Nature. 2002;

156. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, Mori H, et al. TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun. 2006;

157. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, et al. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. 2006;

158. Zheng Q, Huang T, Zhang L, Zhou Y, Luo H, Xu H, et al. Dysregulation of ubiquitin-proteasome system in neurodegenerative diseases. Frontiers in Aging Neuroscience. 2016.

159. Lowe J, Blanchard A, Morrell K, Lennox G, Reynolds L, Billett M, et al. Ubiquitin is a common factor in intermediate filament inclusion bodies of diverse type in man, including those of Parkinson's disease, Pick's disease, and Alzheimer's disease, as well as Rosenthal fibres in cerebellar astrocytomas, cytoplasmic bodies in m. J Pathol. 1988;

160. Cheroni C, Marino M, Tortarolo M, Veglianese P, De Biasi S, Fontana E, et al. Functional alterations of the ubiquitin-proteasome system in motor neurons of a mouse model of familial amyotrophic lateral sclerosis. Hum Mol Genet. 2009;

161. Juarez J, Alatorre-Meda M, Cambon A, Topete A, Barbosa S, Taboada P, et al. Hydration effects on the fibrillation process of a globular protein: the case of human serum albumin. Soft Matter [Internet]. 2012;8(13):3608–19. Available from: http://dx.doi.org/10.1039/C2SM06762E

162. Rubin J, Khosravi H, Bruce KL, Lydon ME, Behrens SH, Chernoff YO, et al. Ion-specific effects on prion nucleation and strain formation. J Biol Chem. 2013/08/31. 2013;288(42):30300–8.

163. Quintana C, Cowley JM, Marhic C. Electron nanodiffraction and high-resolution electron microscopy studies of the structure and composition of physiological and pathological ferritin. J Struct Biol. 2004;

164. Sana B, Poh CL, Lim S. A manganese-ferritin nanocomposite as an

ultrasensitive T2contrast agent. Chem Commun. 2012;

165. Wenborn A, Terry C, Gros N, Joiner S, D'Castro L, Panico S, et al. A novel and rapid method for obtaining high titre intact prion strains from mammalian brain. Sci Rep. 2015;

166. Zhou Z, Fan JB, Zhu HL, Shewmaker F, Yan X, Chen X, et al. Crowded cell-like environment accelerates the nucleation step of amyloidogenic protein misfolding. J Biol Chem. 2009;284(44):30148–58.

167. Terry C, Wenborn A, Gros N, Sells J, Joiner S, Hosszu LLP, et al. Ex vivo mammalian prions are formed of paired double helical prion protein fibrils. Open Biol. 2016;

168. Liu J, Liu Z, Zhang Y, Yin F. A novel antagonistic role of natural compound icariin on neurotoxicity of amyloid beta peptide. Indian J Med Res. 2015/09/12. 2015;142(2):190–5.

169. Sun CS, Lee CC, Li YN, Yao-Chen Yang S, Lin CH, Chang YC, et al. Conformational switch of polyglutamine-expanded huntingtin into benign aggregates leads to neuroprotective effect. Sci Rep. 2015/10/10. 2015;5:14992.

170. Oladzad Abbasabadi A, Javanian A, Nikkhah M, Meratan AA, Ghiasi P, Nemat-Gorgani M. Disruption of mitochondrial membrane integrity induced by amyloid aggregates arising from variants of SOD1. Int J Biol Macromol. 2013/07/23. 2013;61:212–7.

171. Yedlapudi D, Joshi GS, Luo D, Todi S V, Dutta AK. Inhibition of alpha-synuclein aggregation by multifunctional dopamine agonists assessed by a novel in vitro assay and an in vivo Drosophila synucleinopathy model. Sci Rep. 2016/12/06. 2016;6:38510.

172. Ngo ST, Steyn FJ. The interplay between metabolic homeostasis and neurodegeneration: insights into the neurometabolic nature of amyotrophic lateral sclerosis. Cell Regen. 2015/09/01. 2015;4(1):5.

173. Palamiuc L, Schlagowski A, Ngo ST, Vernay A, Dirrig-Grosch S, Henriques A, et al. A metabolic switch toward lipid use in glycolytic muscle is an early pathologic event in a mouse model of amyotrophic lateral sclerosis. EMBO Mol Med. 2015/03/31. 2015;7(5):526–46.

174. Saez I, Vilchez D. The Mechanistic Links Between Proteasome Activity, Aging and Age-related Diseases. Curr Genomics. 2014/03/22. 2014;15(1):38–51.

175. Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron. 2013/08/13. 2013;79(3):416–38.

176. Weedon MN, Hastings R, Caswell R, Xie W, Paszkiewicz K, Antoniadi T, et al. Exome sequencing identifies a DYNC1H1 mutation in a large pedigree with dominant Axonal Charcot-Marie-Tooth disease. Am J Hum Genet. 2011;

177. Tsurusaki Y, Saitoh S, Tomizawa K, Sudo A, Asahina N, Shiraishi H, et al. A DYNC1H1 mutation causes a dominant spinal muscular atrophy with lower extremity predominance. Neurogenetics. 2012;

178. Willemsen MH, Vissers LEL, Willemsen MAAP, van Bon BWM, Kroes T, de Ligt J, et al. Mutations in DYNC1H1 cause severe intellectual disability with neuronal migration defects. J Med Genet. 2012;

179. Shah PR, Ahmad-Annuar A, Ahmadi KR, Russ C, Sapp PC, Horvitz HR, et al. No association of DYNC1H1 with sporadic ALS in a case-control study of a northern European derived population: A tagging SNP approach. Amyotroph Lateral Scler. 2006;

180. El-Kadi AM, Bros-Facer V, Deng W, Philpott A, Stoddart E, Banks G, et al. The Legs at odd angles (Loa) mutation in cytoplasmic dynein ameliorates mitochondrial function in SOD1G93Amouse model for motor neuron disease. J Biol Chem. 2010;

181. Kabashi E, Agar JN, Strong MJ, Durham HD. Impaired proteasome function in sporadic amyotrophic lateral sclerosis. Amyotroph Lateral Scler. 2012/05/29. 2012;13(4):367–71.

182. Plc PS. Proteome Sciences website [Internet]. Available from: https://www.proteomics.com/

183. Eisen a, Schulzer M, MacNeil M, Pant B, Mak E. Duration of amyotrophic lateral sclerosis is age dependent. Muscle Nerve. 1993;

184. Logroscino G, Traynor BJ, Hardiman O, Chió A, Mitchell D, Swingler RJ, et al. Incidence of amyotrophic lateral sclerosis in Europe. J Neurol Neurosurg Psychiatry. 2010;

185. Ting L, Rad R, Gygi SP, Haas W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods. 2011;

186. Database MT human disease. ALS elite genes [Internet]. Available from: https://www.malacards.org/card/amyotrophic_lateral_sclerosis_1#Related Genes-table

187. Atkinson AJ, Colburn WA, DeGruttola VG, DeMets DL, Downing GJ, Hoth DF, et al. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. Clinical Pharmacology and Therapeutics. 2001.

188. World Health Organization, International Programme on Chemical Safety. Biomarkers in risk assessment: validity and validation. Environ Heal. 2001;

189. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;

190. Strimbu K, Tavel J a. What are Biomarkers? Curr Opin HIV AIDS. 2011;

191. Szelechowski M, Amoedo N, Obre E, Léger C, Allard L, Bonneu M, et al. Metabolic Reprogramming in Amyotrophic Lateral Sclerosis. Sci Rep. 2018;

192. Tefera TW, Borges K. Metabolic dysfunctions in amyotrophic lateral sclerosis pathogenesis and potential metabolic treatments. Frontiers in Neuroscience. 2017.

193. Delaye JB, Patin F, Piver E, Bruno C, Vasse M, Vourc'h P, et al. Low IDL-B and high LDL-1 subfraction levels in serum of ALS patients. J Neurol Sci. 2017;

194. Offen D, Halevi S, Orion D, Mosberg R, Stern-Goldberg H, Melamed E, et al. Antibodies from ALS patients inhibit dopamine release mediated by L-type calcium channels. Neurology. 1998;

195. Orozco D, Edbauer D. FUS-mediated alternative splicing in the nervous system: Consequences for ALS and FTLD. Journal of Molecular Medicine. 2013.

196. Chang SH, Hwang CS, Yin JH, Chen S Der, Yang DI. Oncostatin M-dependent Mcl-1 induction mediated by JAK1/2-STAT1/3 and CREB contributes to bioenergetic improvements and protective effects against mitochondrial dysfunction in cortical neurons. Biochim Biophys Acta - Mol Cell Res. 2015;

197. Beers DR, Zhao W, Wang J, Zhang X, Wen S, Neal D, et al. ALS patients' regulatory T lymphocytes are dysfunctional, and correlate with disease progression rate and severity. JCI. 2017;

198. Esko JD, Kimata K, Lindahl U. Chapter 16 Proteoglycans and Sulfated Glycosaminoglycans. In: Essentials of Glycobiology. 2009.

199. Caterson B, Melrose J. Keratan sulfate, a complex glycosaminoglycan with unique functional capability. Glycobiology. 2018;28(4):182–206.

200. Hirano K, Ohgomori T, Kobayashi K, Tanaka F, Matsumoto T, Natori T, et al. Ablation of Keratan Sulfate Accelerates Early Phase Pathogenesis of ALS. PLoS One. 2013;

201. Foyez T, Takeda-Uchimura Y, Ishigaki S, Narentuya N, Zhang Z, Sobue G, et al. Microglial keratan sulfate epitope elicits in central nervous tissues of transgenic model mice and patients with amyotrophic lateral sclerosis. Am J Pathol [Internet]. 2015;185(11):3053–65. Available from: http://dx.doi.org/10.1016/j.ajpath.2015.07.016

202. Sarrazin S, Lamanna WC, Esko JD. Heparan sulfate proteoglycans. Cold Spring Harb Perspect Biol. 2011;3(7):1–33.

203. Nishitsuji K. Heparan sulfate S-domains and extracellular sulfatases (Sulfs): their possible roles in protein aggregation diseases. Glycoconj J. 2018;387–96.

204. Holmes BB, DeVos SL, Kfoury N, Li M, Jacks R, Yanamandra K, et al. Heparan sulfate proteoglycans mediate internalization and propagation of specific proteopathic seeds. Proc Natl Acad Sci. 2013;

205. Ancsin JB. Amyloidogenesis: Historical and modern observations point to heparan sulfate proteoglycans as a major culprit. Amyloid. 2003;10(2):67–

79.

206. Shijo T, Warita H, Suzuki N, Kitajima Y, Ikeda K, Akiyama T, et al. Aberrant astrocytic expression of chondroitin sulfate proteoglycan receptors in a rat model of amyotrophic lateral sclerosis. J Neurosci Res. 2018;96(2):222–33.

207. Forostyak S, Homola A, Turnovcova K, Svitil P, Jendelova P, Sykova E. Intrathecal delivery of mesenchymal stromal cells protects the structure of altered perineuronal nets in SOD1 rats and amends the course of ALS. Stem Cells. 2014;

208. DeWitt DA, Richey PL, Praprotnik D, Silver J, Perry G. Chondroitin sulfate proteoglycans are a common component of neuronal inclusions and astrocytic reaction in neurodegenerative diseases. Brain Res. 1994;656(1):205–9.

209. Sasaki S. Autophagy in spinal cord motor neurons in sporadic amyotrophic lateral sclerosis. J Neuropathol Exp Neurol [Internet]. 2011;70. Available from: https://doi.org/10.1097/NEN.0b013e3182160690

210. Sullivan PM, Zhou X, Robins AM, Paushter DH, Kim D, Smolka MB, et al. The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway. Acta Neuropathol Commun. 2016;

211. Shi Y, Lin S, Staats KA, Li Y, Chang WH, Hung ST, et al. Haploinsufficiency leads to neurodegeneration in C9ORF72 ALS/FTD human induced motor neurons. Nat Med. 2018/02/06. 2018;

212. Liu QY, Lei JX, Sikorska M, Liu R. A novel brain-enriched E3 ubiquitin ligase RNF182 is up regulated in the brains of Alzheimer's patients and targets ATP6V0C for degradation. Mol Neurodegener. 2008;3(1):1–16.

213. Mangieri LR, Mader BJ, Thomas CE, Taylor CA, Luker AM, Tse TE, et al. ATP6V0C knockdown in neuroblastoma cells alters autophagy-lysosome pathway function and metabolism of proteins that accumulate in neurodegenerative disease. PLoS One. 2014;9(4).

214. Walls KC, Ghosh AP, Franklin A V., Klocke BJ, Ballestas M, Shacka JJ, et al. Lysosome dysfunction triggers Atg7-dependent neural apoptosis. J Biol Chem. 2010;285(14):10497–507.

215. Kikuchi H, Yamada T, Furuya H, Doh-ura K, Ohyagi Y, Iwaki T, et al. Involvement of cathepsin B in the motor neuron degeneration of amyotrophic lateral sclerosis. Acta Neuropathol. 2003;

216. Stoka V, Turk V, Turk B. Lysosomal cathepsins and their regulation in aging and neurodegeneration. Ageing Research Reviews. 2016.

217. Bae EJ, Yang NY, Lee C, Kim S, Lee HJ, Lee SJ. Haploinsufficiency of cathepsin D leads to lysosomal dysfunction and promotes cell-to-cell transmission of α-synuclein aggregates. Cell Death Dis. 2015;

218. Morena F, Argentati C, Trotta R, Crispoltoni L, Stabile A, Pistilli A, et al. A

comparison of lysosomal enzymes expression levels in peripheral blood of mild- and severe-Alzheimer's disease and MCI patients: Implications for regenerative medicine approaches. Int J Mol Sci. 2017;

219. Li L, Gao L, Song Y, Qin ZH, Liang Z. Activated cathepsin L is associated with the switch from autophagy to apoptotic death of SH-SY5Y cells exposed to 6-hydroxydopamine. Biochem Biophys Res Commun. 2016;

220. Babcock DT, Shen W, Ganetzky B. A neuroprotective function of nsf1 sustains autophagy and lysosomal trafficking in drosophila. Genetics. 2014;

221. Song P, Trajkovic K, Tsunemi T, Krainc D. Parkin Modulates Endosomal Organization and Function of the Endo-Lysosomal Pathway. J Neurosci. 2016;

222. Ghomashchi F, Naika GS, Bollinger JG, Aloulou A, Lehr M, Leslie CC, et al. Interfacial kinetic and binding properties of mammalian group IVB phospholipase A2 (cPLA2??) and comparison with the other cPLA2 isoforms. J Biol Chem. 2010;

223. Singer AG, Ghomashchi F, Le Calvez C, Bollinger J, Bezzine S, Rouault M, et al. Interfacial kinetic and binding properties of the complete set of human and mouse groups I, II, V, X, and XII secreted phospholipases A2. J Biol Chem. 2002;

224. Prasad SS, Garg A, Agarwal AK. Enzymatic activities of the human AGPAT isoform 3 and isoform 5: localization of AGPAT5 to mitochondria. J Lipid Res. 2011;

225. Cao J, Li JL, Li D, Tobin JF, Gimeno RE. Molecular identification of microsomal acyl-CoA:glycerol-3-phosphate acyltransferase, a key enzyme in de novo triacylglycerol synthesis. Proc Natl Acad Sci U S A. 2006;

226. Shindou H, Shimizu T. Acyl-CoA:lysophospholipid acyltransferases. Journal of Biological Chemistry. 2009.

227. Fantini J, Yahi N. Brain Lipids in Synaptic Function and Neurological Disease. Brain Lipids in Synaptic Function and Neurological Disease. 2015.

228. Stillwell W. Membrane Polar Lipids. In: An Introduction to Biological Membranes. 2016.

229. Texidó L, Hernández S, Martín-Satué M, Povedano M, Casanovas A, Esquerda J, et al. Sera from amyotrophic lateral sclerosis patients induce the non-canonical activation of NMDA receptors "in vitro." Neurochem Int. 2011;

230. Kooijman EE, Chupin V, de Kruijff B, Burger KNJ. Modulation of membrane curvature by phosphatidic acid and lysophosphatidic acid. Traffic. 2003;

231. Blasco H, Veyrat-Durebex C, Bocca C, Patin F, Vourc'H P, Kouassi Nzoughet J, et al. Lipidomics Reveals Cerebrospinal-Fluid Signatures of ALS. Sci Rep. 2017;

232. Supattapone S. Phosphatidylethanolamine as a prion cofactor: Potential implications for disease pathogenesis. Prion. 2012.

233. Vilalta A, Brown GC. Neurophagy, the phagocytosis of live neurons and synapses by glia, contributes to brain development and disease. FEBS J. 2017;

234. Lobasso S, Tanzarella P, Vergara D, Maffia M, Cocco T, Corcelli A. Lipid profiling of parkin-mutant human skin fibroblasts. J Cell Physiol. 2017;

235. Saito-Diaz K, Chen TW, Wang X, Thorne CA, Wallace HA, Page-Mccaw A, et al. The way Wnt works: Components and mechanism. Growth Factors. 2013;

236. Kalderon D. Similarities between the Hedgehog and Wnt signaling pathways. Trends in Cell Biology. 2002.

237. Veeman MT, Axelrod JD, Moon RT. A second canon: Functions and mechanisms of β-catenin-independent Wnt signaling. Developmental Cell. 2003.

238. Sun TQ, Lu B, Feng JJ, Reinhard C, Jan YN, Fantl WJ, et al. PAR-1 is a Dishevelled-associated kinase and a positive regulator of Wnt signalling. Nat Cell Biol. 2001;

239. Yang Y. Wnts and Wing: Wnt Signaling in Vertebrate Limb Development and Musculoskeletal Morphogenesis. Birth Defects Research Part C - Embryo Today: Reviews. 2003.

240. Chen Y, Guan Y, Liu H, Wu X, Yu L, Wang S, et al. Activation of the Wnt/β-catenin signaling pathway is associated with glial proliferation in the adult spinal cord of ALS transgenic mice. Biochem Biophys Res Commun. 2012;

241. Chen Y, Guan Y, Zhang Z, Liu H, Wang S, Yu L, et al. Wnt signaling pathway is involved in the pathogenesis of amyotrophic lateral sclerosis in adult transgenic mice. Neurol Res. 2012;

242. Li X, Guan Y, Chen Y, Zhang C, Shi C, Zhou F, et al. Expression of Wnt5a and its receptor Fzd2 is changed in the spinal cord of adult amyotrophic lateral sclerosis transgenic mice. Int J Clin Exp Pathol. 2013;

243. Lev N, Barhum Y, Lotan I, Steiner I, Offen D. DJ-1 Knockout augments disease severity and shortens survival in a mouse model of ALS. PLoS One. 2015;

244. Knippenberg S, Sipos J, Thau-Habermann N, Körner S, Rath KJ, Dengler R, et al. Altered expression of DJ-1 and PINK1 in sporadic ALS and in the SOD1G93A ALS mouse model. J Neuropathol Exp Neurol. 2013;

245. Salahuddin P, Rabbani G, Khan RH. The role of advanced glycation end products in various types of neurodegenerative disease: a therapeutic approach. Cellular and Molecular Biology Letters. 2014.

246. Emendato A, Milordini G, Zacco E, Sicorello A, Piaz FD, Guerrini R, et al. Glycation affects fibril formation of a peptides. J Biol Chem.

2018;293(34):13100–11.

247. Iannuzzi C, Irace G, Sirangelo I. Differential effects of glycation on protein aggregation and amyloid formation. Front Mol Biosci. 2014/01/01. 2014;1:9.

248. Vicente Miranda H, Outeiro TF. The sour side of neurodegenerative disorders: the effects of protein glycation. J Pathol. 2010/02/27. 2010;221(1):13–25.

249. Zubiri I, Lombardi V, Bremang M, Mitra V, Nardo G, Adiutori R, et al. Tissue-enhanced plasma proteomic analysis for disease stratification in amyotrophic lateral sclerosis. 2018;2:1–17.

250. Shih YT, Hsueh YP. The involvement of endoplasmic reticulum formation and protein synthesis efficiency in VCP- and ATL1-related neurological disorders Julie Y.H. Chan. Journal of Biomedical Science. 2018.

251. Niikura T, Kita Y, Abe Y. SUMO3 modification accelerates the aggregation of ALS-Linked SOD1 mutants. PLoS One. 2014;

252. Foran E, Rosenblum L, Bogush AI, Trotti D. Sumoylation of critical proteins in amyotrophic lateral sclerosis: Emerging pathways of pathogenesis. NeuroMolecular Medicine. 2013.

253. Lamark T, Johansen T. Aggrephagy: Selective disposal of protein aggregates by macroautophagy. International Journal of Cell Biology. 2012.

254. Kim SH, Shanware NP, Bowler MJ, Tibbetts RS. Amyotrophic lateral sclerosis-associated proteins TDP-43 and FUS/TLS function in a common biochemical complex to co-regulate HDAC6 mRNA. J Biol Chem. 2010;

255. Kwiatkowski TJ, Bosco DA, LeClerc AL, Tamrazian E, Vanderburg CR, Russ C, et al. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science (80-). 2009;

256. Vance C, Scotter EL, Nishimura AL, Troakes C, Mitchell JC, Kathe C, et al. ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. Hum Mol Genet. 2013;

257. Polymenidou M, Cleveland DW. Biological spectrum of amyotrophic lateral sclerosis prions. Cold Spring Harb Perspect Med. 2017;

258. Safar JG, Wille H, Geschwind MD, Deering C, Latawiec D, Serban A, et al. Human prions and plasma lipoproteins. Proc Natl Acad Sci. 2006;

259. Chou SM, Taniguchi A, S. Wang H, Festoff BW. Serpin=serine protease-like complexes within neurofilament conglomerates of motoneurons in amyotrophic lateral sclerosis. J Neurol Sci [Internet]. 1998 Oct;160:S73–9. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022510X98002020

260. Cuerrier D, Moldoveanu T, Davies PL. Determination of peptide substrate specificity for??-calpain by a peptide library-based approach: The

importance of primed side interactions. J Biol Chem. 2005;

261. Nixon RA. The calpains in aging and aging-related diseases. Ageing Research Reviews. 2003.

262. Sorimachi H, Mamitsuka H, Ono Y. Understanding the substrate specificity of conventional calpains. Biol Chem. 2012;393(9):853–71.

263. Park JY, Jang SY, Shin YK, Suh DJ, Park HT. Calcium-dependent proteasome activation is required for axonal neurofilament degradation. Neural Regen Res. 2013;

264. Carchon HA, Jansen E, Eggermont E. Determination of enterokinase activity by measuring the disappearance of trypsinogen. 1992;207:175–83.