

## Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

---

Mfon Umoh

---

Date

Characterization of clinical, pathological, and proteomic differences along the Amyotrophic lateral sclerosis and Frontotemporal dementia disease spectrum

By

Mfon Effiong Umoh  
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science  
Neuroscience

---

Jonathan D. Glass, M.D.  
Advisor

---

Gary Bassell, Ph.D.  
Committee Member

---

Marla Gearing, Ph.D.  
Committee Member

---

Lian Li, Ph.D.  
Committee Member

---

Nicholas Seyfried, Ph.D.  
Committee Member

---

Thomas Wichmann, M.D.  
Committee Member

Accepted:

---

Lisa A. Tedesco, Ph.D.  
Dean of the James T. Laney School of Graduate Studies

---

Date

Characterization of clinical, pathological, and proteomic differences  
along the Amyotrophic lateral sclerosis and Frontotemporal dementia  
disease spectrum

By

Mfon Effiong Umoh  
BAS, University of Pennsylvania, 2011

Advisor: Jonathan D. Glass, M.D.

An abstract of  
A dissertation submitted to the Faculty of the  
James T. Laney School of Graduate Studies of Emory University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in Graduate Division of Biomedical and Biological Sciences  
Neuroscience  
2017

## ABSTRACT

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are two neurodegenerative diseases that share genetic, clinical, and pathological overlap. This overlap has been strengthened by the identification of the C9orf72 genetic expansion mutation. This hexanucleotide repeat (GGGGCC) expansion within the Chromosome 9 open reading frame 72 (C9orf72) gene was identified in 2011 and is recognized as the most common genetic cause of both ALS and FTD. The goal of this dissertation project was to increase our understanding of the clinical, pathological, and proteomic differences in patients with and without this expansion mutation as well as to gain a better understanding of the ALS-FTD disease spectrum. Using DNA samples from an entire ALS clinic population spanning 14 years, demographic, clinical, and survival differences were compared in ALS patients with and without the expansion to understand the clinical consequences of the C9orf72 expansion. Post mortem tissue from patients along the clinical spectrum of ALS and FTD were analyzed to gain a better understanding of the disease pathobiology. Specifically, mass spectrometry was used to analyze frontal cortex tissue samples to investigate differential protein expression within the insoluble proteome across the ALS-FTD disease spectrum and relate differences to the hallmark TDP-43 pathology present in patient brains. Finally, mass spectrometry was also used to investigate the total proteome across the clinical continuum and to generate a protein co-expression network that allows us to better understand the pathways implicated across these diseases and the contribution of the C9orf72 genetic mutation. Our results show that there are very few demographic and clinical differences in the C9orf72 ALS population, but there is reduced survival in this group and an increased presence of comorbid frontotemporal dementia when compared to the general ALS population not carrying the C9orf72 mutation. Proteomic analysis showed differences in protein co-

expression along the disease spectrum and a distinct difference in C9orf72 ALS cases compared to sporadic ALS. These differences were validated by comparing our network to externally generated networks and using immunohistochemistry and immunoblotting techniques. Overall, these studies provide important new information regarding the clinical differences conferred by the C9orf72 genetic expansion and the molecular underpinnings of the ALS-FTD disease spectrum. The unbiased proteomic data from these clinical populations deliver new avenues for future investigation, including potential targets for understanding pathogenesis and designing therapeutic interventions.

Characterization of clinical, pathological, and proteomic differences  
along the Amyotrophic lateral sclerosis and Frontotemporal dementia  
disease spectrum

By

Mfon Effiong Umoh  
BAS, University of Pennsylvania, 2011

Advisor: Jonathan D. Glass, M.D.

A dissertation submitted to the Faculty of the  
James T. Laney School of Graduate Studies of Emory University  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy  
in Graduate Division of Biomedical and Biological Sciences  
Neuroscience  
2017

## ACKNOWLEDGEMENTS

I am grateful to the many people who made this work possible. I would first like to acknowledge God for seeing me through this process, without him and Jesus Christ as my personal savior I would not have been able to complete this work.

Next, I would like to thank the many individuals that helped me complete the projects described in this document. Thanks to my mentor Dr. Jonathan Glass, for your mentorship and enthusiasm through the many ups and downs of research. Thank you for allowing me to direct my projects based on my interests and challenging me to bring new expertise to the Lab. Thanks to my committee members for your time and input on my projects. Your feedback was instrumental in shaping this work. A special thanks to Dr. Nicholas Seyfried and Dr. Eric Dammer for the many hours of training you provided for me.

A special thanks to Dr. Marla Gearing for your mentorship and guidance from the start of my graduate school rotations. Thanks to the Glass Lab members, Yingjie Li and Seneshaw Asress, for being great colleagues and keeping my spirits up. Special thanks to Dr. Tezeta Tadesse for being a great role model and lifelong mentor. Your friendship and guidance kept me sane through this whole process. Many thanks also to my other friends for your encouragement through this process, especially the San Diego crew.

Finally, I would like to thank and dedicate this dissertation to my family. My father, Patrick Umoh, and my mother, Grace Ebong, moved my 4 older sisters Akaniyene Udoh, Dr. Uduak Andy, Dr. Edo McGee, and Nse Esema, and I to the United States from Nigeria when I was 5 years old. They left their country and sacrificed all that they knew and loved so that my sisters and I could pursue any dreams we had. Their love and sacrifice has shaped who I am. In a time where immigration is such a heavily debated topic, we are one of many examples of the great

contributions immigrants make to this country. Many thanks to my sisters who have encouraged my curious nature since childhood and continually set the bar high. Ladies, thank you for all you have showed me and continue to teach me. I would also like to acknowledge my grandmother, Arit Ebong, who helped raised me. Her strength, grace, and brilliance have always inspired me. I am eternally grateful to my family for their love and unending support. One of my favorite authors, Junot Diaz writes “we cannot realize that which we can't imagine”. Family, thank you for allowing my imagination to run wild!

## TABLE OF CONTENTS

Abstract .....	IV
Acknowledgements .....	VII
Table of Contents.....	IX
List of Figures .....	XII
List of Tables.....	XIV
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
○ Clinical description of ALS, FTD, and overlap syndrome.....	2
○ Historical overview.....	5
○ Epidemiological overview & prognosis.....	7
○ Pathological overview.....	10
○ Genetic landscape.....	11
○ C9orf72 overview.....	15
○ Mass spectrometry based proteomics for neurodegenerative disease research....	19
○ Proposed research.....	24
○ Research overview.....	25
<b>CHAPTER 2: MATERIALS AND METHODS.....</b>	<b>28</b>
○ Case materials.....	29
▪ Samples for clinical analyses.....	29
▪ Samples for pathological analyses.....	29
▪ Acquisition of clinical and demographic information from database.....	38
○ DNA extraction.....	38
○ Repeat primed PCR for identification of C9orf72 expansion.....	39

○ Family history investigation.....	40
○ Antibodies.....	44
○ Immunocytochemistry.....	44
○ Western blotting.....	45
○ Proteomics sample preparation.....	48
▪ Insoluble homogenate.....	48
▪ Total Homogenate.....	54
○ Label Free Mass spectrometry- Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS).....	55
○ Statistics .....	57
▪ Statistics for clinical analyses.....	57
▪ Statistics for pathological analyses.....	59
▪ Statistics for proteomic analyses.....	59
○ Rationale.....	67
<b>CHAPTER 3: COMPARATIVE ANALYSIS OF C9ORF72 AND SPORADIC DISEASE IN AN ALS CLINIC POPULATION.....</b>	<b>72</b>
• Introduction.....	73
• Results.....	78
• Discussion.....	89
<b>CHAPTER 4: TDP-43 INTERACTOME IN THE INSOLUBLE PROTEOME ALONG THE ALS-FTD DISEASE SPECTRUM .....</b>	<b>97</b>
• Introduction.....	98
• Results.....	103

• Discussion.....	119
<b>CHAPTER 5: INTEGRATED NETWORK APPROACH ACROSS ALS-FTD FRONTAL CORTEX TOTAL HOMOGENATE PROTEOME.....</b>	<b>124</b>
• Introduction.....	125
• Results.....	127
• Discussion.....	156
<b>CHAPTER 6: SUMMARY AND FUTURE DIRECTIONS.....</b>	<b>160</b>
• Summary of the work.....	161
• Therapeutic implications.....	164
• Future directions.....	165
• Conclusions.....	169
<b>REFERENCES.....</b>	<b>170</b>

## LIST OF FIGURES

1.1. Diagram illustrating genetic overlap between ALS and FTD.....	14
1.2. Prominent hypotheses related to C9orf72 expansion mutation.....	18
1.3. Overview of Mass spectrometry based protein identification.....	20
2.1. Electropherograms contrasting C9Pos patient and C9Neg patient.....	41
2.2. Prompt for family history follow-up.....	43
2.3 Extraction protocol used to generate detergent-insoluble fraction.....	49
2.4 Coomassie stained SDS gel of samples for insoluble proteome analysis.....	51
2.5 Spectral count representation of relative protein abundance in LC-MS/MS.....	53
2.6 SDS-PAGE gel for total proteome analysis quality control.....	58
2.7 Proteomic quality control measures and data analysis pipeline.....	62
3.1 Flowchart of family history of Amyotrophic lateral sclerosis (ALS) in screened cohort.....	81
3.2 Classification of familial and sporadic disease in the C9orf72 cohort.....	83
3.3 Survival analysis comparing C9Pos and C9Neg ALS patients.....	84
3.4 Survival analysis restricted to $\leq 10$ years follow-up time.....	85
3.5. Representative images of p62 immunohistochemical analysis of cerebellar sections.....	90
4.1. Significant changes in the detergent-insoluble brain proteome.....	107
4.2. Segregation of clinical groups by principal component analysis of detergent-insoluble LFQ proteomic data.....	111
4.3. TDP-43 levels across samples used for detergent-insoluble proteomic analysis.....	113
4.4. Network analysis of protein co-aggregation based on correlation with TDP-43 in detergent insoluble proteome.....	118
5.1. Segregation of clinical ALS-FTD phenotypes by proteomic signatures .....	133

5.2. Integrated protein co-expression and differential expression across the ALS-FTD disease continuum.....	134
5.3. TDP-43 pathological and proteomic profiles across clinical groups.....	136
5.4. Protein co-expression organizes the proteome into modules associated with brain cell types and clinical and pathological traits.....	138
5.5 Validating the ALS-FTD protein network against previously generated Emory Brain protein network from multiple neurodegenerative diseases.....	140
5.6. Modules with causal links to ALS are associated with inflammation and enriched with TDP-43 protein-protein interactions (PPIs).....	145
5.7. Astrocyte and microglial markers associated with neuroinflammation are increased across the ALS-FTD spectrum.....	150
5.8. Co-expression analysis resolves C9orf72 specific changes in ALS brain.....	152
5.9. C9orf72 protein reduction in C9Pos ALS compared to C9Neg ALS.....	154

## LIST OF TABLES

2.1 Demographic and clinical characteristics comparing C9Pos ALS cases to C9Neg ALS cases.....	31
2.2. Demographic, clinical, and pathological features of autopsy cohort.....	32
2.3. Trait table for cases included in analysis of insoluble proteome.....	34
2.4. Trait table for cases included in analysis of total proteome.....	36
2.5. Antibodies.....	46
2.6 Fisher’s exact test for PPIs enrichment analysis.....	66
3.1. Tabular comparison of studies reporting on clinical and demographic characteristics in C9Pos patients compared to C9Neg patients.....	74
3.2. Demographic and clinical characteristics comparing C9Pos ALS cases to C9Neg ALS cases.....	79
3.3. Cox regression model.....	86
4.1. Relative abundance for all detergent insoluble proteins identified.....	106
4.2. Proteins most correlated with TDP-43(TARDBP) in the detergent -insoluble human brain proteome.....	117
5.1. Differential and co-expression protein analysis.....	132
5.2. Gene ontology (GO) term enrichment for co-expression network .....	132

**Chapter 1: Introduction**

*Clinical overview*

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease that is primarily associated with degeneration of motor neurons in the motor cortex, brainstem, and spinal cord. Frontotemporal dementia (FTD) is a form of dementia associated with degeneration of neurons in the frontal and temporal lobes of the brain. These diseases are so different clinically that patients with either typically present at different clinics. Nevertheless, there is a subset of patients that develop both of these seemingly disparate diseases, and there are some families with individuals with either ALS, FTD, or both diseases.

Clinically, the presentations of these diseases are dramatically different. The primary symptom of ALS is progressive weakness. ALS patients can have limb onset or bulbar onset of disease. Limb onset often leads to loss of ambulatory ability or hand and arm function, while bulbar disease affects speech, eating, and swallowing. Though weakness typically starts in one area, eventually it spreads to other regions to involve all extremities (muscles of the trunk and abdomen) and bulbar face mouth and throat muscles. Muscles that control eye movements and bowel or bladder sphincters are typically spared. Lower motor neuron (LMN) signs, such as weakness and muscle atrophy, are typically accompanied by upper motor neuron (UMN) signs of spasticity and brisk reflexes (gag, jaw, and limb) (Lomen-Hoerth 2011). ALS is a notoriously heterogeneous disease that begins, spreads, and progresses at different rates in different patients. Death from ALS is a consequence of weakness of respiratory muscles, which leads to respiratory failure.

FTD patients present with progressive changes in behavior, personality, language, and loss of executive function due to neurodegeneration in the frontal and temporal lobes of the brain. The main clinical syndromes encompassing FTD are behavioral variant (bvFTD), semantic variant (SD) (Seelaar, Kamphorst et al.), and progressive non-fluent aphasia (PNFA). bvFTD is characterized by progressive behavioral impairment and loss of executive function associated with atrophy predominantly in the frontal lobe. bvFTD starts with noticeable behavioral changes (i.e. apathy or disinhibition), then progresses to a more pronounced and disabling condition that may include more compulsive behaviors and worsened cognitive problems (i.e. planning and attention deficits), and finally reaches profound behavioral issues that may also be associated with memory loss and language difficulty. Clinical disease progression typically follows increased atrophy of frontal lobes and expansion of atrophy into the tips of the temporal lobes and basal ganglia (deeper brain structures involved in motor coordination, cognition, emotion, and learning); this can usually be followed grossly by MRI imaging. The semantic variant is characterized by a loss of object knowledge with anomia (a form of aphasia where patients are unable to recall names of everyday objects) and has predominantly asymmetrical atrophy of the anterior temporal lobes of the brain. Patients with the semantic variant usually begin with deficits in empathy or awareness of other's emotions, then progress to difficulty recognizing names and faces of people they know or understanding other people, and ultimately very poor communication and lots of behavioral issues (i.e. disinhibition, compulsions, apathy). PNFA is characterized by expressive or motor speech deficits due to peri-sylvian atrophy (Rainero, Rubino et al. 2017). Patients with PNFA have trouble speaking and producing language, this usually starts with slowed speech and trouble getting words out, it progresses to the use of shortened sentences that lack articles and adjectives, and these patients ultimately progress are

mute and have behavioral problems. Progression of disease is variable depending on the variant and varies from person to person. FTD patients usually physically decline and end stage FTD patients may even have difficulty swallowing, chewing, and moving their bladder/bowels; death in FTD patients usually occurs due to physical changes and is most common from infections in the lung, skin, or urinary tract.

There are specific criteria used in the diagnosis of these diseases. ALS is diagnosed using the revised El Escorial criteria which are based on the involvement of several regions and the presence of upper and lower motor neuron signs (Brooks, Miller et al. 2000). Diagnostic standards for a clinical diagnosis of FTD are based on the Neary criteria, which include behavioral changes, language difficulties, and executive function problems (Neary, Snowden et al. 1998).

Traditionally, people with FTD were not thought to experience motor dysfunction, and people with ALS were expected to be spared from symptoms related to cognitive domains. It is now increasingly recognized, however, that a significant proportion of patients experience symptoms of both conditions. Indeed, a quarter of ALS patients have a comorbid diagnosis of FTD, and frontal dysfunction, not reaching the full extent of clinical FTD, has been identified in up to 50% of ALS patients (Lomen-Hoerth, Anderson et al. 2002, Lomen-Hoerth, Murphy et al. 2003). Even in these patients where behavioral changes do not qualify for a diagnosis of FTD, the changes and disturbances for themselves and their families can be noticeable and difficult (Murphy, Henry et al. 2007). Similarly, 15% of FTD patients develop comorbid ALS (Lomen-Hoerth 2011). Additionally, there are FTD patients who have symptoms of motor neuron disease

but are not diagnosed with ALS. bvFTD is the most common form of dementia seen concurrently with ALS. Thus, this overlap syndrome is seen in patients presenting with either ALS or FTD, and can also be seen as a combination of ALS and FTD that present simultaneously (Lomen-Hoerth, Anderson et al. 2002).

### *Historical perspective*

The first description of ALS is attributed to Jean-Marie Charcot, a French neurologist who is considered by some to be the father of neurology, instrumental in explaining how the central nervous system functions. His account of ALS was published in 1869 and was based on his careful observation of the clinical and pathological features of patients with isolated motor syndromes (Goetz 2000, Ferrari, Kapogiannis et al. 2011). He named the disease based on the clinical and pathological picture presented by his patients; “amyotrophic” denotes the loss of muscle mass, while “lateral sclerosis” refers to the scarring of the lateral corticospinal tracts of the spinal cord. By linking symptoms and clinical manifestations of the disease to known functions of certain brain and spinal cord regions, Charcot defined the disease quite accurately. He even noted that non-motor areas of the brain may be involved in onset or progression of the disease (Geser, Lee et al. 2010).

FTD was first described by Arnold Pick in 1892 in a patient with progressive aphasia and anterior lobar atrophy. This patient had bouts of aggressiveness and showed progressive cognitive impairment (Berrios and Girling 1994). At autopsy this patient’s brain showed atrophy of the temporal lobe of the left hemisphere (Ferrari, Kapogiannis et al. 2011). Other case reports and neuropathological analysis of postmortem brains from patients with this clinical presentation

revealed argyrophilic neuronal inclusions, termed Pick bodies (Dickson 1998). After descriptions of this clinical presentation circulated, it became increasingly recognized and is now presumed to be the second most common form of early onset dementia in patients under age 65. Importantly, FTD is associated with a variety of pathologies and Pick bodies are relatively rare, though this anecdotal discovery highlights the utility of neuropathological and clinical correlations in better understanding disease.

The overlap of ALS and FTD has been noted for more than a century, and is acknowledged by contemporary neurology (Ziegler 1930, Finlayson and Martin 1973, Hudson 1981). Mitsuyama and colleagues described a large cohort of 71 patients with dementia and motor neuron disease (Mitsuyama 1993). Neary and colleagues added to this by studying a smaller cohort with the overlap, describing changes based on imaging and neuropathological analysis of a small number of these cases (Neary, Snowden et al. 1990). Awareness of distinct neuropathology increased with the identification of ubiquitin immunoreactive inclusions in the cytoplasm of motor neurons and similar inclusions in the extra-motor cortex in ALS patients and ALS patients with dementia, respectively (Leigh, Anderton et al. 1988, Leigh, Whitwell et al. 1991). The link between ALS and FTD was further strengthened in the early 2000s by clinical descriptions by Lomen-Hoerth and colleagues (Lomen-Hoerth, Anderson et al. 2002, Lomen-Hoerth, Murphy et al. 2003). In 2006, the landmark identification of TDP-43 as the major protein in the ubiquitinated inclusion pathology of both ALS and FTD patient brains further strengthened the link between these diseases (Neumann, Sampathu et al. 2006). Further evidence connecting these conditions came most recently in 2011 with the identification of the C9orf72 genetic mutation in families with patients with either clinical phenotype or individuals with the overlap syndrome.

Several famous individuals throughout history have been affected by these devastating diseases. Lou Gehrig, a famous Yankee baseball player was diagnosed with ALS at age 36 and died 2 years later; ALS is commonly referred to as Lou Gehrig's disease in North America. The famous physicist, Stephen Hawking, was diagnosed with ALS at age 21 and, as evidence to the heterogeneity of the disease, is currently 75 years old. Famous individuals diagnosed with FTD also include Baltimore Colts tight-end football player and football hall of fame inductee John Mackey, who was diagnosed at age 65 and died at age 69. Terry Jones, the Monty Python star actor and movie director was diagnosed with FTD and is reported to have the primary progressive aphasia variant. The phenotypes represented by these individuals demonstrates the wide range of patients seen along this disease spectrum.

### *Epidemiology and prognosis*

The number of patients nationwide and worldwide with diseases along the ALS-FTD spectrum is not insignificant, and with such an impactful diagnosis the need to better understand the pathogenesis and progression of these diseases is great. In the United states approximately 5,000 people are diagnosed yearly with ALS (Chio, Logroscino et al. 2013). Most newly diagnosed individuals are between ages 45 and 65, though earlier onset and later onset forms do exist (Mehta, Antao et al. 2014). Epidemiological studies show a worldwide incidence of 2-3 cases of ALS per year per 100,000 population over the age of 15, and a lifetime risk of ALS of 1:350 for men and 1:400 for women (Chio, Logroscino et al. 2013). Currently there are 30,000 people in the US living with ALS. The cumulative time-dependent survival in ALS at 1, 5, and 10 years from diagnosis is 76.2%, 23.4%, and 11.8%, respectively (Pupillo, Messina et al. 2014). The

median survival time from onset of symptoms to death ranges from 20 to 36 months (del Aguila, Longstreth et al. 2003, Millul, Beghi et al. 2005, Chio, Logroscino et al. 2009). Hereditary (familial) forms of ALS account for 10% of all patients. Men have a higher incidence of disease (3.0 per 100,000 person-years, 95% CI:2.8-3.3 compared to 2.4 per 100,000 person-years, 95% CI: 1.2-2.6 in women) (Kiernan, Vucic et al. 2011), though the incidence between men and women is equal for those with familial forms of the disease. Interestingly, though aging is a major factor, the incidence of ALS decreases rapidly after age 80 (Kiernan, Vucic et al. 2011). There is very little known about racial and ethnic differences in the incidence of ALS but it is presumed, based on limited studies, that there are no racial or ethnic differences.

Limited numbers of FTD studies have estimated the prevalence of FTD at 15-22 per 100,000 people within age 45 to 62 years (Ratnavalli, Brayne et al. 2002). FTD is typically seen in people under age 65, and represents the second most common form of dementia within this age groups. Survival in FTD patients is estimated to be 6 to 11 years from disease onset. The behavioral variant form of FTD (bvFTD) is associated with the shortest survival (Hodges, Davies et al. 2003). FTD is more rapidly progressive than Alzheimer's disease (AD), the most common form of dementia (Roberson, Hesse et al. 2005).

The prevalence of the overlap of ALS in FTD has not been precisely determined; it is presumed to be about 15% of patients meeting ALS diagnostic criteria (Lomen-Hoerth, Anderson et al. 2002, Lomen-Hoerth 2011). Patients with both diseases have a reduced estimated survival time (Hu, Shelnutt et al. 2013). In a study conducted by Hu and colleagues, the ALS Brief Cognitive Assessment (ALS-BCA) was conducted in a longitudinal cohort of 171 ALS patients to

determine cognitive predictors of survival. People with both ALS and dementia had shorter overall survival (Hu, Shelnett et al. 2013). The reduced survival was completely accounted for by the presence of disinhibited or apathetic behaviors in the model, whereas demented patients without behavioral changes did not show reduced survival compared to the non-demented group. Prior to that study it was shown that there was roughly a 6-year difference in survival in ALSFTD patients compared to FTD patients (Hodges, Davies et al. 2003, Roberson, Hesse et al. 2005).

The prevalence of FTD in ALS is slightly greater than that of ALS in FTD, reports have ranged from 22% to 48%. Differences in these prevalence estimates are mostly dependent on how FTD is characterized by certain practitioners (Lomen-Hoerth, Murphy et al. 2003, Murphy, Henry et al. 2007).

Risk factors for developing ALS include increasing age, which peaks around age 75 (Ingre, Roos et al. 2015), and a family history of ALS and other neurodegenerative conditions including FTD. Heritability studies suggest that heritable genetics contribute 60% to the risk of developing ALS, while the remaining 40% is thought to be environmental (Al-Chalabi, Fang et al. 2010, Wingo, Cutler et al. 2011, Al-Chalabi and Hardiman 2013). Other purported risk factors include military service, environmental toxins, and cyanobacteria, but these have yet to be proven.

Associated risk factors for FTD include head injury and family history of FTD (Rosso, Landweer et al. 2003, Weder, Aziz et al. 2007). In a case control study that included 80 sporadic FTD cases investigators found a significant association between FTD and head trauma. Several risk factors

have been suggested for dementia in ALS based on data from relatively small cohorts. These include increased age, male gender, low education level, family history of dementia, low forced vital capacity, pseudobulbar palsy and bulbar site of onset (Massman, Sims et al. 1996, Abrahams, Goldstein et al. 1997, Lomen-Hoerth, Murphy et al. 2003, Ng, Rademakers et al. 2015).

### *Pathological overlap*

The histopathological characteristics of frontotemporal lobar degeneration (FTLD, the pathology underlying most FTD cases) and ALS include an abnormal accumulation of protein aggregates in affected regions of the brain, similar to many other neurodegenerative diseases. Although the molecular pathology of FTD is heterogeneous, a significant proportion of cases have ubiquitin-positive inclusions similar to those observed in degenerating neurons in ALS brains. The identification of transactive response DNA binding protein 43kDa (TDP-43) as a major pathological component of these ubiquitinated inclusions in ALS and FTD further connected these diseases with known clinical overlap (Neumann, Sampathu et al. 2006, Sampathu, Neumann et al. 2006). 97% of ALS patients demonstrate neuronal cytoplasmic aggregates of TDP-43 aggregates in the spinal cord and brain, and about half of FTD patients show this same pathology in the frontal and temporal cortices, as well as other brain regions, termed FTLD-TDP (Arai, Hasegawa et al. 2006). FTLD-TDP is the primary pathology underlying the overlap between the clinical syndromes of FTD and ALS (Braak, Ludolph et al. 2010, Ng, Rademakers et al. 2015). TDP-43 in humans is a 414-amino acid RNA binding protein encoded by the TARDBP gene that is ubiquitously expressed and is normally localized to the nucleus (Buratti

and Baralle 2010). In disease, TDP-43 is cleaved, abnormally phosphorylated, and accumulates in the cytoplasm as TDP-43 aggregates.

### *Genetic landscape and overlap*

Findings from genetic studies add further evidence of the overlap of these disparate conditions. 10% of ALS patients have a hereditary form of the disease, while 40% of patients with bvFTD patients are thought to have a familial form of disease (McKhann, Albert et al. 2001, Seelaar, Kamphorst et al. 2008). In patients with the combination of ALS and FTD, familial disease accounts for between 40 and 60% of cases (van der Zee, Gijssels et al. 2013). Within an individual pedigree, family members can display phenotypes across the ALS-FTD disease spectrum. Genetic overlap between ALS and FTD has been recognized for quite some time. There are genetic alterations that have been identified in individuals that have both ALS and FTD; other genetic alterations lead to clinical phenotypes considered purer forms of either disease.

Several disease-causing genetic mutations have been identified in ALS. The first mutation identified was in the superoxide dismutase 1 (SOD-1) gene; SOD-1 mutations account for a small percentage (2%) of all ALS cases (Rosen, Siddique et al. 1993). To date, there are no SOD-1 mutations that have been substantially implicated in FTD. Mutations in Optineurin (OPTN), Fused in sarcoma (FUS), Angiogenin (Ang), Ubiquilin-2 (UBQL2) and vesicle-associated membrane protein associated protein B (VAPB) also account for rare cases of familial ALS (Nishimura, Mitne-Neto et al. 2004, Greenway, Andersen et al. 2006, Blair, Williams et al.

2010, Maruyama, Morino et al. 2010, Deng, Chen et al. 2011). There are other reported disease-causing genes (Taylor, Brown et al. 2016).

In FTD, mutations in the microtubule-associated protein tau (MAPT) and progranulin (PRGN) account for about 25% of the familial cases (Goedert, Crowther et al. 1998, Mackenzie, Baker et al. 2006, van Swieten and Heutink 2008). MAPT mutations lead to Tau pathology and not TDP-43 pathology. Interestingly, though mutations in the TARDBP gene account for a small percent (4%) of familial ALS, no mutations in this gene have been linked to FTD cases even though patients with FTD show the same pathological TDP-43 inclusions seen in ALS (Kabashi, Valdmanis et al. 2008, Sreedharan, Blair et al. 2008). Moreover, though UBQLN2 mutations were identified in a large family with ALS phenotypes, a quarter of the individuals with the mutation had behavioral and executive dysfunction abnormalities similar to FTD (Deng, Chen et al. 2011).

The identification of genetic alterations associated with ALS and FTD have influenced our understanding of these neurodegenerative diseases. Most recently, in 2011, two independent research groups identified Chromosome 9 open reading frame 72 (*C9orf72*) as the major gene causing familial forms of ALS and FTD (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011). It is suggested that *C9orf72* mutations account roughly for 34% of familial ALS cases, 6% of sporadic ALS cases, 26% of familial FTD cases, 5% of sporadic FTD cases, and 0.17% of control subjects (van Blitterswijk, DeJesus-Hernandez et al. 2012). Prior to 2011, a chromosome 9 locus was identified in familial ALS patients with FTD, though the specific gene implicated remained unknown (Morita, Al-Chalabi et al. 2006, Vance, Al-Chalabi et al. 2006, Valdmanis, Dupre et al. 2007). The diagram in **Figure 1.1**, published in a review by

Radford and colleagues, shows the spectrum of disease from ALS to ALSFTD, to FTD along the horizontal axis. On the vertical axis, the diagram shows pathways and potential pathologies associated with genetic alterations (C9orf72 shown in the middle) that have been associated with this spectrum of diseases (Radford, Morsch et al. 2015).

The discovery of the C9orf72 hexanucleotide repeat expansion mutation has accelerated research activity into the overlap between ALS and FTD. In chapter 3 of this dissertation, I explore the relationship between this genetic mutation and clinical presentation, progression and survival in ALS. In chapter 5, I report the use of post-mortem brain samples from a cohort of patients with this genetic mutation to get a clearer picture of the ALS-FTD disease continuum. Studying genetic forms of neurodegenerative diseases, though not representative of all patients, can help us advance our understanding of the pathobiology of these conditions. By integrating clinical and genetic information with pathological and molecular information we get a more comprehensive picture of these diseases.

	ALS		ALSFTD		FTD	
Vascular	ANG					
DNA/RNA		TARDBP, FUS	MATR3			
Toxic protein	SOD1		C9orf72			
Cytoskeleton			TUBA4A		MAPT	
Phagocytosis	PFN1					
Inflammation			SQSTM1, VCP, TBK1	UBQLN2	CHMPB2	TREM2
Protein Degradation	OPTN					GRN

**Figure 1.1. Diagram illustrating genetic overlap between amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD).** Genetic mutations identified in patients are shown linked to clinical phenotypes along the ALS (blue) and FTD (red) spectrum. Genes are plotted based on putative mechanisms in relation to disease (top to bottom). ANG=Angiogenin, SOD1=Superoxide Dismutase-1, PFN1=Profilin 1, OPTN=Optineurin, TARDBP=TAR-DNA Binding Protein, FUS=Fused in Sarcoma, MATR3=Matrin-3, TUBA4A=Tubulin alpha 4A chain, SQSTM1=Sequestome-1, VCP=Valosin Containing Protein, TBK1=TANK- Binding Kinase 1, C9orf72=Chromosome 9 open reading frame 72, CHMPB2=Charged Multivesicular Body Protein 2B, MAPT= Map Tau, TREM2= Triggering Receptor Expressed on myeloid cells 2, GRN= Granulin. *Image is modified and reproduced from Radford et al. Frontiers in cellular neuroscience (2015) under the terms of a creative commons license.*

### *C9orf72 expansion mutation*

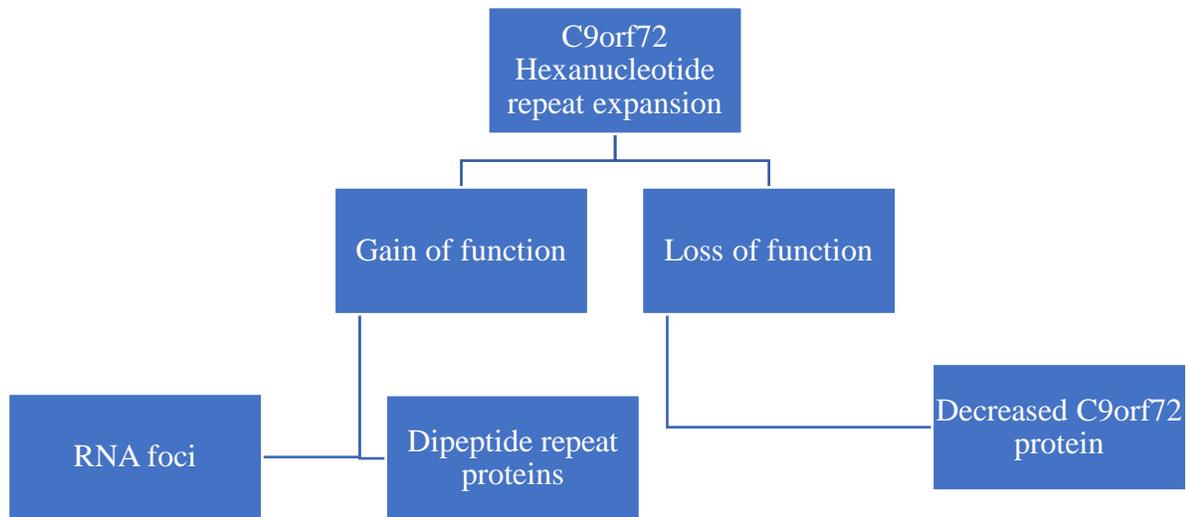
The most notable genetic link between ALS and FTD is the C9orf72 gene. Fewer than 23 repeats of the hexanucleotide GGGGCC in this gene locus corresponds to a wild type allele and more than 30 repeats corresponds to an expanded allele present in disease individuals (Achi and Rudnicki 2012). The initial identification of the C9orf72 expansion found a prevalence of the expansion in 12% of familial FTD and 22.5% of familial ALS cases (DeJesus-Hernandez, Mackenzie et al. 2011), while a simultaneous report found higher prevalence rates with 46% in familial ALS, 21% in sporadic ALS, and 29% in familial FTD (Renton, Majounie et al. 2011).

After this mutation was initially identified, patients with the C9orf72 expansion were thought to differ clinically from typical ALS and FTD patients. Specifically, a predominance of psychiatric symptoms was reported in some cohorts. These symptoms ranged from delusions, irrational thoughts, and paranoia (Snowden, Rollinson et al. 2012). Chapter 3 of this dissertation investigates the clinical differences in C9orf72 ALS compared to sporadic ALS in a clinical cohort. Additionally, C9orf72 patients were thought to have distinct neuropathological features compared with ALS and FTD patients without the C9orf72 expansion mutation. Specifically, overwhelming p62-positive and TDP-43 negative cytoplasmic inclusions in the hippocampus and cerebellar regions of the brain were reported to be a hallmark pathology for C9orf72 cases (Boeve, Boylan et al. 2012, Cooper-Knock, Hewitt et al. 2012). Similar to differences seen in reports from clinical cohorts comparing C9orf72 cases to sporadic cases, there were differences in reports from different autopsy cohorts (Murray, DeJesus-Hernandez et al. 2011, Snowden, Rollinson et al. 2012). Chapter 3 of this dissertation investigates neuropathological differences in a small autopsy cohort, comparing C9orf72 ALS/FTD cases to sporadic ALS/FTD cases. Though

the cerebellum is a region not typically implicated in ALS and FTD, several studies have tried to improve our understanding of cerebellar involvement based on the regional neuropathological differences present in C9orf72 cases; this demonstrates how neuropathological findings can inform clinical surveillance practices.

Currently there are four major hypotheses regarding how the C9orf72 genetic mutation leads to disease. These hypotheses are outlined in **Figure 1.2**, taken from a recent review article (Taylor, Brown et al. 2016). The first is focused on loss of C9orf72 function, thought to occur due to decreased C9orf72 mRNA expression. Reduced C9orf72 protein levels in C9ALS/FTD patients support the theory that C9orf72 haploinsufficiency is involved in disease pathogenesis. Loss of C9orf72 function has been proposed to lead to an inhibition of autophagy and a promotion of neuroinflammation as these processes have been linked to C9orf72 function (Waite, Baumer et al. 2014). Recent studies investigating the function of the C9orf72 protein have shown that WDR41 (WD Repeat-Containing Protein 41) interacts with the C9orf72/SMCR8 heterodimer and this is associated with a complex essential for autophagy initiation (FIP200/ULK1 complex) (Sullivan, Zhou et al. 2016). A second hypothesis on C9orf72 related pathogenesis is that sense and antisense repeat-containing RNA generate toxic RNA foci, which are thought to sequester RNA binding proteins. This sequestration creates a sink of RNA binding proteins leading to abnormal RNA metabolism and function. Sense and antisense RNA foci transcribed from the C9orf72 repeat expansion are detected in the CNS and spinal fluid of C9Pos patients (DeJesus-Hernandez, Mackenzie et al. 2011, Mizielinska, Lashley et al. 2013, Gendron, Chew et al. 2017) and several studies have linked abnormal RNA function to C9orf72 disease pathobiology (Lee, Chen et al. 2013, Mori, Lammich et al. 2013, Xu, Poidevin et al. 2013, Cooper-Knock, Walsh et

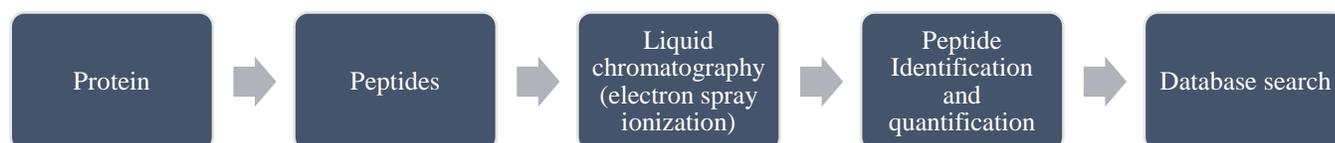
al. 2014, Conlon, Lu et al. 2016). A third theory regarding C9orf72 pathogenesis is that toxicity is conferred by sense and antisense repeat containing transcripts which go through a noncanonical form of translation termed repeat-associated non-ATG (RAN) translation (Zu, Liu et al. 2013). Sense and antisense transcripts of the hexanucleotide repeat ( $G_4C_2$ ) go through RAN translation in all reading frames, generating dipeptide repeat proteins (DPRs: GP, GA, GR, PA, and PR). Aggregates of these DPRs have been shown to be toxic in several models of disease (Kwon, Xiang et al. 2014, May, Hornburg et al. 2014, Mizielinska, Gronke et al. 2014, Wen, Tan et al. 2014, Tran, Almeida et al. 2015, Lopez-Gonzalez, Lu et al. 2016, Ohki, Wenninger-Weinzierl et al. 2017). Finally, another hypothesized mechanism of C9orf72 disease pathogenesis is disruption of nucleocytoplasmic transport through nuclear pores due to the presence of RNA foci, DPRs, or direct toxicity to nucleopore proteins (Freibaum, Lu et al. 2015, Zhang, Donnelly et al. 2015, Zhang, Gendron et al. 2016). These potential mechanisms may not be mutually exclusive and further studies on C9orf72 patients and disease models may improve our understanding of disease pathogenesis in C9orf72-related disease, and possibly even in nongenetic forms of these diseases. One method to investigate the role of the C9orf72 expansion in the ALS-FTD disease spectrum that has not been explored by many groups is mass spectrometry based proteomics.



**Figure 1.2. Prominent hypotheses related to C9orf72 expansion mutation.** The C9orf72 expansion mutation results in a modest reduction in the levels of the C9orf72 protein (left) that might contribute to disease progression through loss of function mechanisms such as abnormal microglial responses. The expression of sense and antisense RNA transcripts that contain the expanded repeat may drive a toxic gain of function (right). The two prominent gain-of-function modes that are implicated are: toxicity through the sequestration of RNA-binding proteins in RNA foci by the expanded GGGGCC repeat RNA transcript; and the production of DPR proteins through RAN translation, leading to toxicity through several cellular targets such as membrane-less organelles and nuclear pores.

### *Mass spectrometry*

Mass spectrometry, because of its high sensitivity and high-throughput nature, is a useful tool for identifying and accurately quantifying proteins within a complex sample. Advances in mass spectrometry-based proteomics, mostly related to tools (instrumentation, software, data collection strategies, and analysis), have allowed us the ability to identify thousands of proteins in a single experiment. The underlying technology behind mass spectrometry involves chemical fragmentation of a sample into charged particles (ions) and the subsequent measurement of charge and mass of the resulting particles (done by passing the particle through electric and magnetic fields). Shotgun proteomics is used in the systematic identification of proteins from complex samples. This approach involves liquid chromatography (LC) for the separation of peptides produced by trypsin digestion and their subsequent analysis by tandem mass spectrometry (MS/MS). The overview of this process is diagramed in **Figure 1.3** (Gstaiger and Aebersold 2009).



**Figure 1.3. Overview of Mass spectrometry based protein identification.** Homogenized brain samples are digested with trypsin and the resulting peptide mixtures are fractionated using reverse-phase liquid chromatography (LC). The fractionated peptide solution is then exposed to an electric potential, this causes a spray to form which leads to the desolvation and ionization of the peptides (electrospray ionization). Mass to charge ( $m/z$ ) ratios are measured from peptide ions that pass the collision cell without fragmentation in the mass spectrometer (MS). Specific ions are randomly selected for collision-induced dissociation with neutral gas (i.e. helium) and the resulting fragment ions are measured in the second mass analyzer in tandem mass spectrometry (MS/MS). The MS precursor ion intensities obtained in the first stage can be used for peptide quantification, while the MS/MS fragment ion information from the second stage has sequence information that can be compared from *in silico* digested protein sequence databases that can then be used for peptide and subsequent protein identification.

The tools used for analysis have a critical effect on the mass accuracy, resolving power, sensitivity and dynamic range of peptide identification and quantification. Briefly, a mass spectrometer consists of three important parts: the ion source which transforms the molecules into ionized fragments, the mass analyzer which sorts the ions based on their mass to charge ( $m/z$ ) ratio, and a detector which measures the ion intensity and thus gives information about abundance of specific fragments. The *tandem aspect* of mass spectrometry (MS/MS) allows for sequencing of peptides by using the fragmentation pattern generated by peptides. Within the mass analyzer a specific peptide is selected, its ions are stabilized in the collision cell and collide with an inert gas which causes them to fragment by collision induced dissociation. The resulting peptide fragments then get sorted in a second mass analyzer and get recorded by a detector which creates a tandem mass spectrum. This spectrum can later be analyzed by comparing it within a database of tandem mass spectra that lead to peptide identification which can in turn be used to identify its protein of origin. Proteomics allows for the identification of changes in protein expression and pathways with high precision, accuracy and reproducibility (Farrelly, Föcking et al. 2014). Proteomics is the only technique currently available that is used to systematically characterize molecular alterations at the protein level.

We used a systems approach to analyze mass spectrometry data generated from the total proteome using frontal cortex samples. This serves as an alternative to investigating individual protein alterations, as is done with differential expression analysis, and moves towards a method of assessing protein networks that may be altered in ALS and FTD. We believe this approach possibly provides an advantage of looking at the whole range of pathways that may be involved in complex neurodegenerative diseases. This approach differs from much research in basic and

translational neuroscience, which is focused on candidate gene identification and candidate hypotheses (Parikshak, Gandal et al. 2015). The greatest value in the unbiased proteomic approach is that we are able to look at, and measure, entire systems of proteins. Generally, science tackles large questions by simplifying them to smaller pieces. However, biology is inherently complex and this approach may affect our understanding of complex systems, especially the brain. Our systems approach using weighted correlation network analysis to better understand the ALS-FTD disease spectrum allows for multiple pathways to be altered at the same time and commences with no a priori assumptions.

Handling biological complexity using a systems approach is essential to uncover the underlying mechanisms of neurodegenerative disorders, especially ALS and FTD. Properly designed and reproducible molecular profiling studies allow for the simultaneous appraisal of hypotheses in an impartial manner and the creation of new hypotheses. Networks can be large and seemingly complex, but they can provide an organizational structure that streamlines testing, and generates hypotheses, and ultimately identifies therapeutic targets. The overall model of employing correlational molecular networks to understand molecular systems changes can be applied across methodologies and enables the investigation of relationships that span multiple levels of analysis (Parikshak, Gandal et al. 2015). Co-expression networks provide a contextual biological framework for discovery and hypothesis driven research with the ultimate goal of emphasizing uniting features of suspected disease proteins, which pushes our understanding of networks implicated in disease (Gaiteri, Ding et al. 2014).

The projects described throughout this dissertation are focused on better understanding clinical, pathological, and proteomic differences in people with ALS and FTD, including those with and without the C9orf72 mutation. A big question in the field that has gone unanswered is why do these seemingly disparate diseases occur together? The results of the studies described here add to our understanding of the co-occurrence of these diseases. We uncover differences and similarities in patients with and without the C9orf72 mutation expansion that move the field forward by revealing networks altered along the ALS-FTD continuum, this increases of our understanding of the disease pathobiology. Delineating clinical differences in C9orf72 ALS patients, compared to sporadic ALS patients, contributes a valuable resource on a significantly sized north American cohort. Investigating the detergent-insoluble brain proteome of patients along this spectrum and correlating differences in the proteome to pathological burden adds to our understanding of the involvement of TDP-43 in disease pathogenesis across this disease spectrum. Co-expression network analysis of the total brain proteome uncovered expected and novel alterations in the brains of patients along the ALS-FTD diseases spectrum. The inclusion of a cohort of patients with the C9orf72 expansion also identified novel proteomic differences in the brains of this patient population.

## **Proposed research**

The goals of this dissertation are to characterize clinical, pathological, and proteomic differences in patients along the ALS-FTD disease spectrum. We wish to investigate clinical features in C9orf72 expansion positive patients compared to sporadic patients. We also hope to identify proteins and biological pathways within the frontal cortex that are centrally involved in the ALS-FTD disease spectrum. To address these goals, we will compare demographic, clinical, and survival characteristics in C9orf72 ALS patients to sporadic ALS (sALS) patients using a substantial clinical cohort. We will employ label free mass spectrometry based proteomics to examine proteomic differences across clinically characterized cases along the ALS-FTD disease spectrum, including a subset of cases with the C9orf72 expansion mutation. Through probing the detergent-insoluble brain proteome as well as the total brain proteome, we will correlate pathological features present in patient brains to other protein changes that may inform our understanding of affected pathways that converge or diverge based on clinical phenotypes. Using an integrated network approach, we hope to define and validate differences in the total proteome from patient brains with different clinical phenotypes and genetic propensities along this spectrum of diseases to create a systems level picture of this disease spectrum. Our use of a quantitative proteomics approaches is essential for the creation of an unbiased and systematic profile of proteins, altered through differential expression or co-expression, representing the ALS-FTD disease spectrum.

We believe that patients with the C9orf72 expansion mutation represent a separate ALS population with clinical differences and survival differences compared to sporadic ALS patients. We hypothesize that altered proteins in the brains of patients along the ALS-FTD disease

spectrum mediate the pathogenesis and progression of these diseases. Thus, their identification, characterization, and validation will enhance our understanding of mechanisms underlying these complex diseases. This work may also help guide the development of targeted therapies for patients within the ALS-FTD disease spectrum by revealing unique and converging pathway alterations. It is our hope that the studies described in this dissertation will serve to advance scientific understanding of these devastating disorders and in some way influence our clinical practice as it relates to presentation and progression of ALS in C9orf72 patients.

### **Research overview**

**Aim 1: Investigate whether the C9orf72 expansion mutation in patients with ALS is associated with unique demographic and clinical features.**

Previous studies in C9orf72 ALS patients (C9Pos) showed a more aggressive and rapidly progressive disease when compared to sporadic patients. These conclusions were based on a limited number of studies on small cohorts that reported conflicting results. Some showed a higher prevalence of bulbar onset, earlier age at onset, reduced survival, and a higher incidence of co-morbid dementia when compared to patients with sporadic disease (Murray, DeJesus-Hernandez et al. 2011, Chio, Borghero et al. 2012, Stewart, Rutherford et al. 2012, Irwin, McMillan et al. 2013, Snowden, Harris et al. 2013). Other studies describe C9Pos ALS patients as exhibiting clinical features similar to those with sporadic disease, only with a more rapidly progressive disease course (Cooper-Knock, Hewitt et al. 2012, Sabatelli, Conforti et al. 2012). Thus, it was unclear whether patients with the expansion (C9Pos) are phenotypically different from patients without the expansion (C9Neg). We tested this using DNA from patients attending the Emory ALS clinic between 2001 and 2015 and clinical information from the research

database (Umoh, Fournier et al. 2016). This research cohort of 781 patients was screened for the C9orf72 expansion, and demographic and clinical data were compared between those with and without the C9orf72 mutation. For mutation carriers without a family history of ALS, family history of dementia and other non-ALS neurodegenerative diseases in first-degree relatives was assessed. The C9orf72 expansion was identified in 61 patients (7.8%). Compared to those without the expansion mutation, these patients did not differ in race, age, or site of onset. C9orf72 patients were more likely to have a family history of ALS (59% vs 7.9%) and to present with comorbid frontotemporal dementia (FTD) (14.8% vs 1.7%). Survival was significantly shorter in patients with the expansion. Subsequent investigation of patients initially categorized as having no known family history of ALS identified a family history of dementia in 16 cases and showed that 6 of these had characteristics suggestive of FTD.

**Aim 2: Identify proteomic targets selectively altered along the ALS-FTD disease spectrum and validate candidate proteomic targets and networks.**

Post mortem frontal cortex tissue selected from patients along the ALS-FTD disease clinical spectrum were analyzed by unbiased, label-free mass spectrometry. Two distinct proteomes were examined; the detergent-insoluble proteome and the total proteome. Differential expression analysis and correlational analysis were used to assess the insoluble proteome data. Weighted network co-expression analysis was applied to the total proteome data and network alterations across the ALS-FTD spectrum were validated using biochemical techniques and through the integration of other 'omic' datasets. From the insoluble proteome work, our findings highlight the utility of label-free mass spectrometry based proteomics to identify disease-specific proteins that are common across TDP-43 proteinopathies and those that are unique to cases with

dementia. Using differential expression analysis, we get a better sense of proteins altered in each disease group, and can begin to infer pathways that may be involved in disease pathogenesis. Analysis and validation of the total proteome identified inflammation associated markers in the frontal cortex of ALSFTD patients compared to patients with only ALS; this finding is consistent with other evidence that the dementia phenotype represents changes within the brain that correlate with, if not also caused by, increased neuroinflammation (Bettcher and Kramer 2013). Our analysis also showed increased expression of inflammatory proteins in C9orf72 ALS patient brains, further implicating C9orf72 in inflammatory pathways. Differences in C9orf72 ALS cases were also validated against previously published RNA-seq data from an independent cohort. Several pathways in addition to brain inflammation are implicated in our dataset, including protein transport, RNA binding, and structural components. This work provides a resource that will invigorate studies moving towards a broad and comprehensive understanding of pathways regulated through biological mechanisms that manifest in co-expression networks with distinguishing characteristics along the ALS-FTD spectrum.

## **Chapter 2: Materials and methods**

## *2.1 Case materials*

### *Samples for clinical analyses*

From January 2001 until December 2015, all patients with a diagnosis of ALS based on a clinical phenotype of progressive upper and lower motor neuron disease, corroborated by exclusionary testing and EMG finding, were recruited for collection of DNA and clinical information for research used in analyses described in chapter 3. A total of 859 blood samples were collected, representing 50% of all eligible patients. The relative proportion of patients contributing DNA was uniformly distributed throughout the 15-year period. **Table 2.1** lists summary characteristics for subjects used in work presented in chapter 3.

### *Post mortem samples for pathological and proteomic analyses*

All post mortem brain tissues were obtained from the Emory Alzheimer's Disease Research Center (ADRC) Brain Bank. Human postmortem tissues were acquired under proper Institutional Review Board (IRB) protocols with consent from family. DNA used in chapter 3 analyses were obtained by extraction from post mortem brain samples for cases where blood was not available (n=18, 2%). Neuropathological analysis comparing C9orf72 positive cases to sporadic cases was conducted on a small autopsy cohort from cases listed in **Table 2.2**. Post mortem tissue samples used in both proteomic studies were selected based on clinical diagnoses. FTD cases were also selected based on neuropathological diagnoses to exclude FTD subtypes that did not have neuropathological diagnosis of FTLD-TDP (frontotemporal lobar degeneration characterized by ubiquitin and TDP-43 positive, tau negative, FUS negative inclusions) pathology. Diagnoses were made by trained neurologists using established clinical criteria (Brooks 1994, Brooks, Miller et al. 2000, McKhann, Albert et al. 2001, Rascovsky, Hodges et al. 2011). Standard

diagnostic neuropathological analysis was performed for all cases, including phosphorylated TDP-43 (pTDP-43) immunohistochemistry on paraffin-embedded tissue sections using the phosphorylated ser409/410 antibody (Cosmo Bio). Ordinal scales were used to assess pTDP-43 pathology (0 to 3) with higher scores indicating greater pathology. The presence of a C9orf72 repeat expansion for samples used was assessed from blood samples using the published repeat primed PCR method (DeJesus-Hernandez, Mackenzie et al. 2011). Clinical and pathological information from all cases, including disease status, neuropathological criteria, age, sex, and post-mortem interval are provided in **Table 2.3** for the insoluble proteome study and **Table 2.4** for the total proteome study.

	<b>C9Pos</b>	<b>C9Neg</b>
<b>N</b>	61	720
<b>Sex (M/F)</b>	30/31	457/263
<b>Age onset</b>	57.9 ± 8.7	56.3 ± 13.5
<b>Site onset bulbar</b>	15 (25%)	191(26%)
<b>Table 2.1. Demographic characteristics for cases used for clinical comparisons.</b> Table lists characteristics of cases included in analysis of clinical cohort described in chapter 3		

<b>Subject</b>	<b>Race/Sex</b>	<b>C9orf72 status</b>	<b>Neuropathological Diagnosis (primary)</b>	<b>Age at onset</b>	<b>Age at death</b>	<b>Disease duration</b>
1	wf	Positive	ALS	51	55	4
2	wf	Negative	ALS	55	59	3.5
3	wm	Negative	ALS	36	43	7
4	wm	Positive	ALS	66	69	3.5
5	wm	Positive	ALS	63	65	2
6	wf	Negative	ALS	69	74	5
7	wm	Negative	ALS	56	64	8
8	bf	Positive	FTLD-TDP	54	57	2.5
9	wm	Negative	Control	NA	57	NA
10	wm	Negative	Control	NA	70	NA
11	wf	Negative	Control	NA	78	NA
12	bf	Negative	Control	NA	43	NA
13	wf	Positive	FTLD-TDP	68	69	<1
14	wf	Negative	FTLD-TDP	66-69	73	4 to 7
15	wm	Negative	FTLD-TDP	56	61	5
16	wm	Positive	FTLD-TDP	79	83	4
17	wf	Negative	FTLD-TDP	62	71	9
18	wm	Positive	FTLD-TDP	57	66	9
19	wm	Negative	FTLD-TDP	57	61	4
20	wf	Positive	FTD/ALS	unavailable	64	Unavailable

**Table 2.2. Demographic, clinical, and pathological features of autopsy cohort.** Table lists characteristics of cases included in neuropathological analysis described in chapter 3. Race (w=white, b=black), Sex (m=male, f-female).

Case ID	Clinical group	CTL path	ALS path	FTD path	PMI (hrs)	Duration (yrs)	Sex	pTDPPrating
als01	1	0	1	0	10	10	1	0
als02	1	0	1	0	11	2	1	0
als03	1	0	1	0	11	2.5	0	0
als04	1	0	1	0	8	3	1	0
als05	1	0	1	0	22	1.5	0	1
als06	1	0	1	0	6	3.5	0	3
ctl01	0	1	0	0	5.5	0	1	0
ctl02	0	1	0	0	20	0	1	0
ctl03	0	1	0	0	35.5	0	1	0
ctl04	0	1	0	0	6	0	0	0
ctl05	0	1	0	0	15.5	0	0	1
ctl06	0	1	0	0	11.5	0	0	0
ftd01	3	0	0	1	9	6	1	3
ftd02	3	0	0	1	11.5	1	1	0
ftd03	3	0	0	1	18	9	0	2
ftd04	3	0	0	1	11.5	10	0	1
ftd05	3	0	0	1	17.5	5	1	2
ftd06	3	0	0	1	6	8	0	3
ftdals01	2	0	1	1	23	7	1	1
ftdals02	2	0	1	1	3	5	0	1
ftdals03	2	0	1	1	32	3	0	1
ftdals04	2	0	1	1	8	7	1	3
ftdals05	2	0	1	1	4	5.5	0	2
ftdals06	2	0	1	1	8.5	1	1	3

**Table 2.3. Trait table for cases included in analysis of insoluble proteome. Clinical and**

pathological traits for cases used in analyses described in chapter 4 are listed. Clinical group (0=ctl, ALS=1, ALSFTD=2, FTD=3), Presence of pathology is indicated in columns listed CTL path, ALS path, FTLD path, based on autopsy analysis, 0 is absent and 1 is present. Post mortem interval (PMI) is listed in hours; duration (in years) is time from disease onset to death. Sex (0=female, 1=male). pTDP rating is on a scale from 0-3 representing amount of phosphorylated TDP pathology in frontal cortex section.

Case ID	Clinical group	PMI (hrs)	Disease duration (yrs)	Sex	C9orf72 status	pTDP rating
CTL1	0	5.5	0	1	0	0
CTL2	0	15.5	0	0	0	1
CTL3	0	6	0	0	0	0
CTL4	0	20	0	1	0	0
CTL5	0	6.5	0	0	0	0
CTL6	0	3	0	0	0	0
CTL7	0	11.5	0	0	0	0
CTL8	0	6.5	0	1	0	0
CTL9	0	35.5	0	1	0	0
CTL10	0	10	0	1	0	0
ALS1	1	6	2.5	0	0	0
ALS2	1	17.5	5	0	1	0
ALS3	1	7.5	5	1	0	0
ALS4	1	2.5	1.5	1	1	3
ALS5	1	11	2	1	0	0
ALS6	1	11	2.5	0	0	0
ALS7	1	14	8	1	0	0
ALS8	1	6	3.5	0	0	3
ALS9	1	6.5	2.5	0	1	1
ALS10	1	5.5	3	0	0	0
ALS11	1	15	4	0	1	0
ALS12	1	10.5	3.5	1	1	0
ALS13	1	7	2	1	1	1
ALS14	1	8	3	1	0	0
ALS15	1	20.5	2.5	0	1	3
ALS16	1	10	10	1	0	0
ALS17	1	6	6	0	0	0
ALS18	1	22	1.5	0	0	1
ALS19	1	15.5	7	0	1	1
ALSFTD1	2	32	3	0	0	1
ALSFTD2	2	8	7	1	0	3
ALSFTD3	2	4	5.5	0	0	2
ALSFTD4	2	18	4.5	1	0	0
ALSFTD5	2	6	9	1	1	0
ALSFTD6	2	8.5	1	1	0	3
ALSFTD7	2	23	7	1	0	1
ALSFTD8	2	9	6	1	0	0
ALSFTD9	2	3	5	0	0	1
ALSFTD10	2	17.5	4	1	0	0

FTD1	3	11.5	10	0	0	1
FTD2	3	17.5	5	1	0	2
FTD3	3	10.5	4	1	1	1
FTD4	3	11.5	5	1	0	2
FTD5	3	5	2	0	0	1
FTD6	3	11.5	1	0	1	0
FTD7	3	6	8	0	0	3
FTD8	3	21	7	1	0	0
FTD9	3	11.5	10	0	0	1
FTD10	3	18	9	0	0	2
FTD11	3	9	6	1	0	3
FTD12	3	17	7.5	1	0	3
FTD13	3	11.5	1	1	0	0

**Table 2.4. Trait table for cases included in analysis of total proteome.** Clinical and pathological traits for cases used in analyses described in chapter 5 are listed. Clinical grouping (0=ctl, ALS=1, ALSFTD=2, FTD=3). Post mortem interval(PMI) is listed in hours; duration (in years) is time from disease onset to death. Sex (0=female, 1=male). pTDP rating is on a scale from 0-3 representing amount of phosphorylated TDP pathology in frontal cortex section.

### *Acquisition of clinical and demographic information from database*

The Emory ALS Center maintains a large collection of DNA samples from patients with ALS along with their demographic and clinical characteristics (Traxinger, Kelly et al. 2013). Demographic (sex, race, and family history of a first-degree relative with ALS) and clinical information (site and age at onset and diagnosis) were obtained at the initial visit. Onset of ALS is defined by the first recognition by the patient of weakness or spasticity. A comorbid diagnosis of FTD was determined by the treating ALS neurologist. Because of the evolving consensus FTD diagnostic criteria throughout the study period, operationalized criteria proposed in 2001 were consistently applied in subsequent years. Information regarding death and tracheostomy-free survival was provided by family members, caregivers, and through searches of public records, including published obituaries and the social security death index (<http://FamilySearch.org>).

### *2.2 DNA extraction*

For analyses described in chapter 3, most (n=763; 98%) DNA came from blood samples that were collected at a clinic visit and processed for DNA extraction according to the instructions provided by the manufacturer of the DNA extraction kit (Qiagen, Venlo, Netherlands) and stored at -80° C until analysis. For samples where DNA was not available from blood DNA was extracted from post mortem brain tissue. The extraction of the DNA from brain samples was achieved in the following manner: 50-100mg of brain tissue was pulverized in lysis buffer and then heated for 15-60 minutes at 65° C. Next, 40ul of proteinase K was added and tubes with samples were inverted repetitively (25X) and incubated in a shaker at 56° C overnight. The following day, additional (20ul) proteinase K was added in and digestion continued for 2-3

hours. After 15ul of RNase A solution was added to the samples, they were inverted repetitively (25X), and incubated at 37° C for 15-60 minutes. Tubes were then rapidly cooled on ice for 3 minutes. Protein precipitation solution was then added to the samples and they were vortexed vigorously for 20s at high speed before centrifugation for 10 minutes at 2000g. The precipitated protein material formed a tight pellet. The supernatant was mixed with 3ml of isopropanol in a new tube by inverting repetitively (50X). This was then centrifuged for 3 minutes at 2000g, after which the supernatant was discarded and the tube was drained by inverting it onto a clean piece of absorbent paper. Then 3ml of 70% ethanol was added to the sample pellet and it was inverted several times. Tubes were then centrifuged for 1 minute at 2000g, the supernatant discarded, and pellets allowed to air dry for 5-10 minutes. Then 400ul of DNA hydration solution was added to the pellets and this was mixed by vortexing for 5 s at medium speed. This mixture was incubated at 65° C for 1 hour to dissolve the DNA. DNA was then incubated with gentle shaking at room temperature overnight. The next day samples were quickly centrifuged, transferred into storage tubes and stored at -80C until analysis. DNA concentrations were measured using a nanodrop (Thermo scientific) and an  $A_{260}/A_{280}$  ratio was evaluated to ensure purity (a ratio between 1.7 and 1.9 was considered to indicate that the DNA was pure).

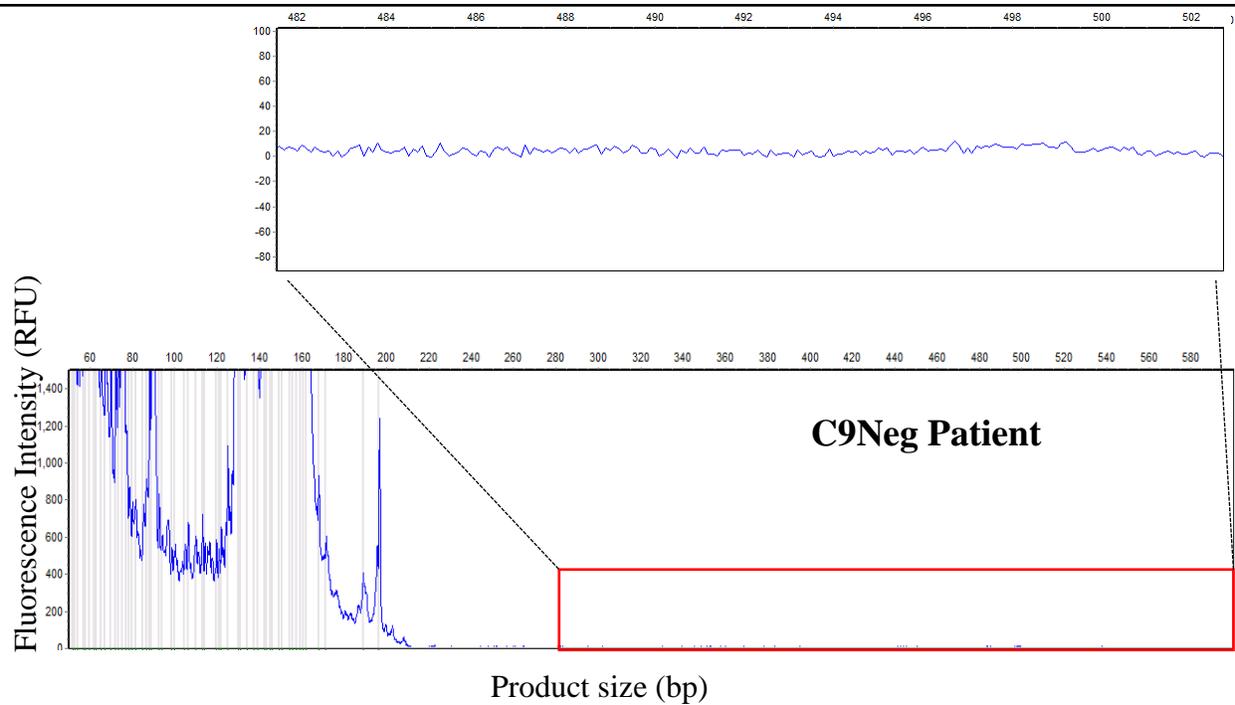
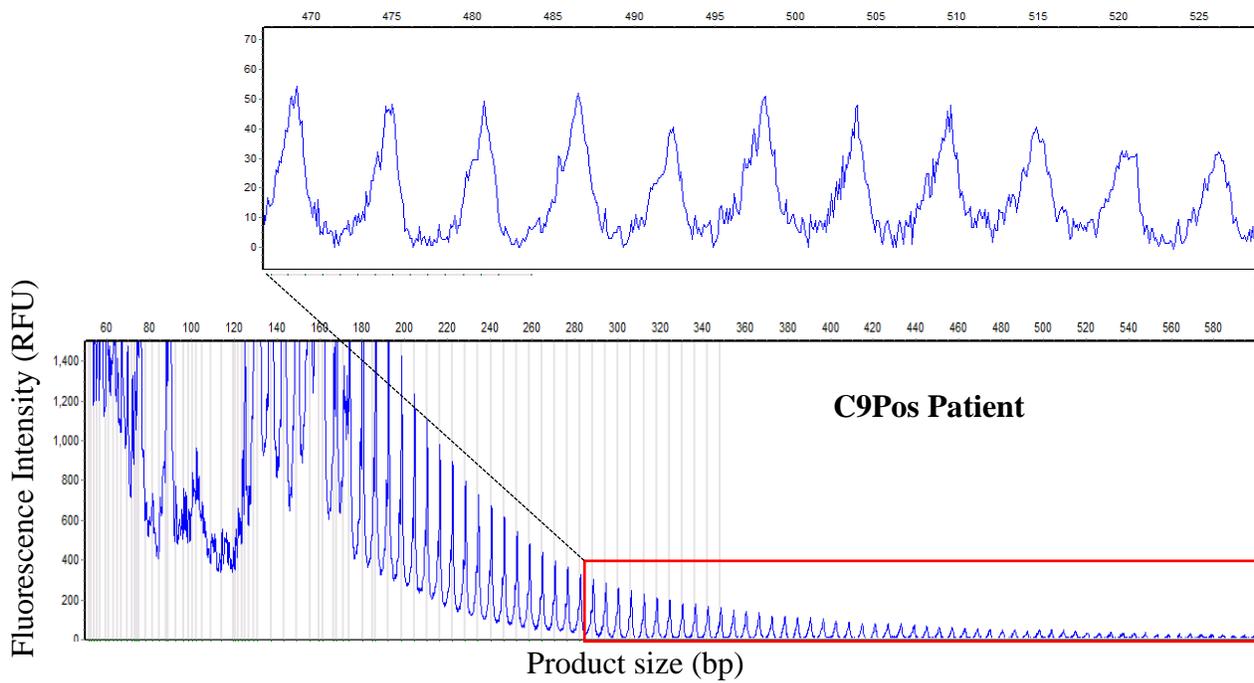
### *2.3 Repeat primed PCR for identification of C9orf72 expansion*

Genotyping for the C9orf72 hexanucleotide repeat was performed using the repeat primed PCR protocol (DeJesus-Hernandez, Mackenzie et al. 2011). DNA was PCR amplified using 3 primers, one of which is fluorescently labeled and incorporates into the amplified product, which is read using capillary electrophoresis on a DNA analyzer (ABI3730; Thermo Fisher, Waltham, MA). Results are interpreted using amplified fragment length polymorphism analysis in GeneMarker

software (Softgenetics, State College, PA); a cutoff of 30 repeats was used to differentiate C9Pos from C9Neg cases. DNA from a C9Pos control from Coriell Institute for Medical Research (6769B1) was included in sample runs. The repeat primed PCR protocol uses a locus-specific flanking primer as well as a paired repeat primer that amplifies from numerous sites within the repeat which generated a characteristic ladder of fragments after capillary electrophoresis. C9Pos and C9Neg electropherograms are shown in **Figure 2.1** to illustrate the difference. Q solution (betaine) was included (final concentration 0.8-1.6M) to improve the amplification of DNA by reducing formation of secondary structures in GC-rich regions; this has been reported to improve the specificity of PCR by eliminating the base pair composition dependence of DNA melting (Frackman, Kobs et al. 1998, Guido, Starostina et al. 2016).

#### *2.4 Family history investigation*

Family history was obtained through a structured interview during clinic visits and recorded on a standardized form. Patients who had a first-degree relative with ALS were categorized as familial ALS; those without a family member with ALS were considered sporadic. To verify the number of sporadic cases in our C9Pos cohort, we telephoned family members (spouses or children) of C9Pos patients without a known family history of ALS. During these follow-up calls, we confirmed the accuracy of the absence of ALS in the family history, and asked about diagnoses of other neurodegenerative diseases in first-degree relatives. If dementia was present in a first-degree relative, we asked about behavioral changes, since behavioral variant FTD is the most common form of dementia seen in conjunction with ALS, regardless of whether the C9orf72 expansion mutation is present. **Figure 2.2** displays follow up prompt used for family history investigation.



**Figure 2.1. Electropherogram demonstrating difference between C9Pos patient and C9Neg patient.** PCR products of repeat-primed PCR reactions are separated on an ABI3730 DNA Analyzer and visualized by GENEMARKER software. The x-axis represents product size (bp) and the y-axis represents fluorescence intensity in relative fluorescence units. Red box on electropherograms are zoomed in to show stutter amplification in C9Pos case and the lack of this pattern in C9Neg case.

Name of ALS patient: \_\_\_\_\_

Date of Birth for ALS patient: \_\_\_\_\_

Who is completing the survey?  Self/ patient with ALS  family member  caregiver

If the survey is being completed by someone other than the patient with ALS, describe your relationship to the patient (example: spouse, mother, son, etc.) \_\_\_\_\_

If the survey is being completed by someone other than the patient with ALS, has the patient passed away?  YES  NO

If yes, when (month/date/year)?

\_\_\_\_\_

**Please answer all questions on behalf of the patient with ALS.**

- Is there any family history of motor neuron disease (ALS)?
- Is there any family history of dementia?
- Is there any family history of neurodegenerative disease?
- Was there any clinically noted changes in cognitive ability in this patient?
- Were there any behavioral issues noted in this patient?

**Figure 2.2. Prompt used for family history follow-up investigation**

## 2.5 Antibodies

Numerous antibodies were used throughout for the work reported in this dissertation to investigate neuropathology present in post mortem samples or validate proteomic differences using immunohistochemistry to visualize pathological differences or immunoblotting to assess differences in protein expression. Commercially available primary antibodies used in these studies are listed in **Table 2.5**. Secondary antibodies used in these studies are also listed in **Table 2.5**.

## 2.6 Immunohistochemistry

Staining was performed using postmortem frontal cortex, hippocampus, and cerebellar tissue for different facets of this work. Human paraffin embedded sections (8 $\mu$ m thickness) were deparaffinized by incubation at 60°C for 30 mins and rehydrated by immersion in xylene and 100% ethanol and 95% ethanol solutions. Antigen retrieval was performed by microwaving slides in 10mM citrate buffer pH 6.0 for 5 mins and then allowing slides to cool to RT for 30 mins. Peroxidase quenching was performed by incubating slides in a 3% hydrogen peroxide solution in methanol for 5 mins at 40°C, Slides were then rinsed in Tris-Brij buffer (1M Tris-Cl pH 7.5, 100mM NaCl, 5mM MgCl<sub>2</sub>, 0.125% Brij 35). For blocking, sections were incubated in normal goat serum or normal horse serum (Elite Vectastain ABC kit), depending on the primary antibody species, for 15 mins at 40°C. Sections were then incubated with primary antibodies (**Table 2.5**) (diluted in 1% BSA in tris-brij 7.5) overnight at 4°C. The following day sections were incubated in biotinylated secondary antibody at 5 $\mu$ L/mL (Elite IgG Vectastain ABC kit) for 30 mins at 37°C and then incubated with the avidin-biotin enzyme complex (Vector Laboratories) for 30 mins. Stains were visualized by incubation with DAB Chromogen (Sigma-

Aldrich) for 5 mins at RT. Slides were counterstained with hematoxylin and then dehydrated in an ethanol series and mounted with cover slips. Slides were analyzed using an Olympus BX51 microscope and imaged with an Olympus DP70 camera.

### *2.7 Western blotting*

Immunoblotting was performed following standard protocols. Total brain homogenates in Laemmli sample buffer were resolved by SDS-PAGE (NuPAGE Bis-Tris (Life Technologies)). Gels were transferred using the iBlot 7-minute blotting system (Thermo Fisher Scientific) dry transfer systems onto nitrocellulose membranes (Invitrogen). Blots were blocked with TBS starting block buffer (Thermo Fisher Scientific) for 30 minutes at room temperature, then probed with primary antibodies (**Table 2.5**) diluted in 10% blocking buffer in PBS overnight at 4C. The following day blots were rinsed and incubated with secondary antibodies conjugated to fluorophores, Alexa Fluor680 goat anti-mouse IgG(H+L) or Alexa Fluor680 goat anti- Rabbit IgG(H+L) (Life Technologies), for one hour at room temperature. Images were captured using an Odyssey Infrared Imaging system (LiCor Biosciences). Blots were quantified using LICOR image studio lite version 5.2.

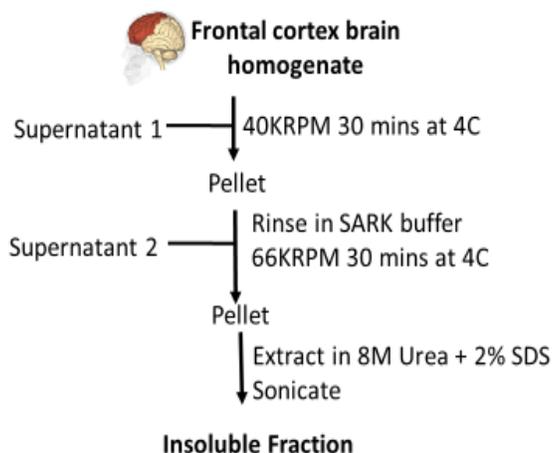
Primary antibodies					
Antibody	Epitope	Host/ Type	Application	Working dilution	Company/ Catalogue #
pTDP-phosphorylated TDP	Phospho Serine 409/410	Rabbit Polyclonal	IHC	1:1000	CosmoBio TIP-PTD-P02
p62	Human p62 Ick ligand aa. 257-437	Mouse Monoclonal	IHC	1:200	BD Biosciences 610833
ionized calcium binding adaptor molecule 1 (Iba-1)	c-terminus of Iba1	Rabbit Polyclonal	IHC	1:500	Wako 019-19741
Glial fibrillary acidic protein (GFAP)	N/A	Mouse Monoclonal	IHC, WB	1:1000	Millipore MAB360
hepatic and glial cell adhesion molecule (HEPACAM)	residues 50-150 of human HEPACAM	Rabbit Polyclonal	IHC, WB	1:1000(Burberry, Suzuki et al.), 5ug/ml (IHC)	Abcam, ab130769
tripeptidyl peptidase 1 (TPP1)	full-length human TPP1 protein	Mouse Polyclonal	IHC, WB	3ug/ml	Sigma Aldrich, WH0001300M1
Moesin (MSN)	N/A	Mouse Monoclonal	IHC, WB	1:2500(Burberry, Suzuki et al.), 2ug/ml(IHC)	Abcam, ab50007
Synaptophysin (SYP)	N/A	Mouse Monoclonal	WB	1:500	Millipore MAB5258
glyceraldehyde 3-phosphate dehydrogenase (GAPDH)	N/A	Mouse Monoclonal	WB	1:1000	Novusbio NB300-221
alpha-tubulin	alpha Tubulin fusion protein ag1727	Rabbit Polyclonal	WB	1:1000	Protein Tech, 11224-1-AP
Secondary antibodies					

Biotinylated Goat anti- Rabbit	Rabbit IgG (H+L)	Goat Polyclonal	IHC	1:1000	Millipore
Biotinylated Goat anti- Mouse	Mouse IgG (H+L)	Goat Polyclonal	IHC	1:1000	Millipore
Alexa Fluor 680 Goat anti- mouse	Mouse IgG (H+L)	Goat Polyclonal	WB	1:10,000	Life technologies
Alexa Fluor 680 Goat anti- Rabbit	Rabbit IgG (H+L)	Goat Polyclonal	WB	1:10,000	Life technologies

**Table 2.5. Antibodies.** Antibodies used for immunohistochemical analysis and for immunoblotting are listed. IHC= immunohistochemistry, WB = western blot.

### *2.8 Insoluble proteome sample preparation: homogenization and proteolytic digestion*

All tissue samples were taken from the prefrontal cortex, specifically the middle frontal gyrus; this region corresponds to Brodmann area 9. This region was selected because of the clinical overlap of cognitive and/or behavioral impairment seen between ALS and FTD patients and because this is a region that has previously been identified to have pathology specific to FTD. The enrichment approach for the detergent-insoluble fraction was conducted as previously described (Seyfried, Gozal et al. 2012). Frozen postmortem tissue (0.5g) was dounce homogenized in 5ml/g (20% w/v) of ice-cold homogenization buffer (50mM HEPES pH7.0, 250 mM sucrose, 1mM EDTA and 1X HALT (Pierce) protease inhibitor cocktail). After samples were homogenized, sarkosyl (N-lauroylsarcosine) and NaCl were added to final concentrations of 1% w/v and 0.5M, respectively (sarkosyl buffer). This initial fraction was defined as the total brain homogenate and was sonicated 3X (Sonic Dismembrator, Fisher Scientific) with 5 sec pulses at 30% amplitude using a microtip probe to shear nucleic acids. To generate sarkosyl-insoluble fractions, total brain homogenates were centrifuged (TFA 100.4 Beckman rotor used) at 180,000 x g for 30 minutes at 4C. The supernatant was retained as the sarkosyl-soluble fractions. The initial pellet was rinsed by resuspending it in 1ml of sarkosyl-buffer and centrifuged at 180,000 x g for an additional 30 minutes. The final pellet was solubilized in urea buffer (8M urea and 2% SDS in 50mM Tris-HCL pH 8.5) to generate the sarkosyl-insoluble fraction. **Figure 2.3** shows the protocol used to generate insoluble fraction samples. Protein concentrations were determined using the bicinchonic acid (BCA) method (Pierce).

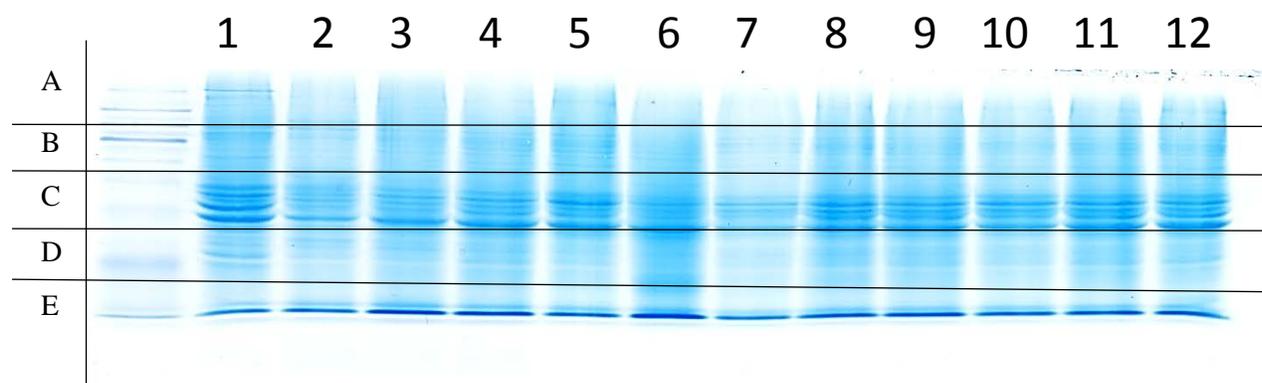
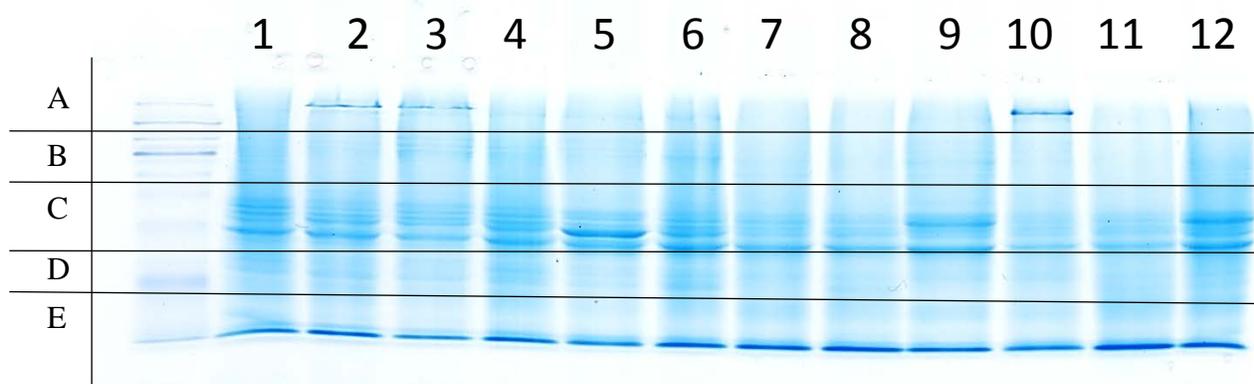


**Figure 2.3. Diagram of extraction protocol used to generate detergent-insoluble fraction for LC-MS/MS analysis described in chapter 4.** Frozen postmortem tissue (0.5g) was dounce homogenized in 5ml/g (20% w/v) of ice-cold homogenization buffer (50mM HEPES pH7.0, 250 mM sucrose, 1mM EDTA and 1X HALT (pierce) protease inhibitor cocktail). After samples were homogenized, sarkosyl (N-lauroylsarcosine) and NaCl were added to final concentrations of 1% w/v and 0.5M, respectively (sarkosyl buffer). This initial fraction was defined as the total brain homogenate and was sonicated 3X (Sonic Dismembrator, Fisher Scientific) with 5 sec pulses at 30% amplitude using a microtip probe to shear nucleic acids. To generate sarkosyl-insoluble fractions, total brain homogenates were centrifuged (TFA 100.4 Beckman rotor used) at 180,000 x g for 30 minutes at 4C. The supernatant was retained as the sarkosyl-soluble fractions. The initial pellet was rinsed by resuspending it in 1ml of sarkosyl-buffer and centrifuged at 180,000 x g for an additional 30 minutes. The final pellet was solubilized in urea buffer (8M urea and 2% SDS in 50mM Tris-HCL pH 8.5).

LC-MS/MS analysis and label free quantification were performed as described previously (Dammer, Duong et al. 2013, Dammer, Lee et al. 2015). The detergent insoluble fraction (20 µg)

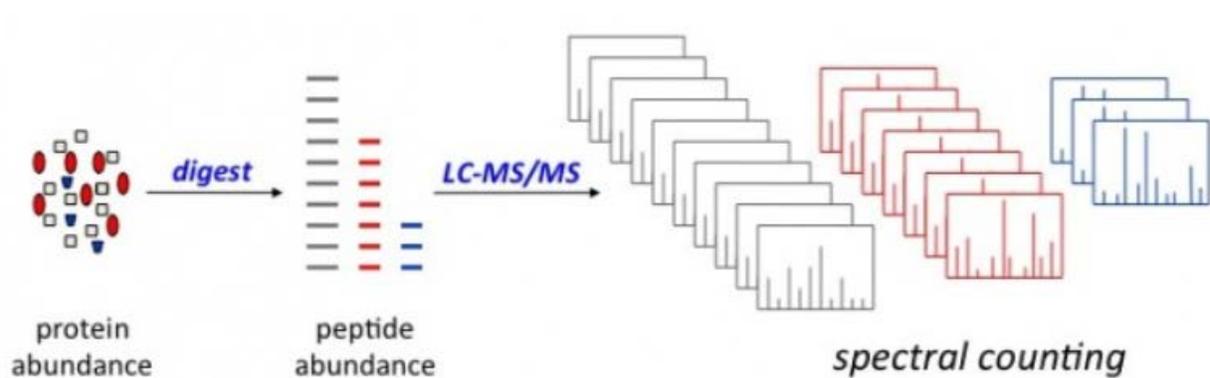
of each case was reduced with 5 mM dithiothreitol (DTT) for 20 minutes at 37°C and then alkylated with 50 mM iodoacetamide (Engelhart, Geerlings et al.) for 30 minutes at 37°C in the dark on a shaker. Alkylated samples were separated on a 10% SDS gel and stained with Coomassie Blue G-250. Each sample lane was cut into five gel bands corresponding to molecular weight ranges to increase the depth of coverage of the proteome (**Figure 2.4**). Gel pieces were then digested overnight with trypsin (12.5 µg/ml) at 37°C. Samples were extracted in a solution of 5% formic acid and 50% acetonitrile (Xiao, MacNair et al.) and following speed-vacuum evaporation an equal amount of each peptide sample was resuspended in loading buffer (0.1% formic acid, 0.03% trifluoroacetic acid, and 1% acetonitrile). The resulting peptides were loaded onto a 20 cm nano-LC column (internal diameter 100 µm) packed with Repronil-Pur 120 C18-AQ 1.9 µm beads and eluted over a 1 h reverse phase gradient comprised of 4-80% buffer B (Buffer A: 0.1% formic acid, 1% acetonitrile in water; Buffer B: 0.1% formic acid in acetonitrile) generated by a NanoAcquity UPLC system (Waters Corporation). Peptides were ionized with electrospray ionization (2.0 kV) and detected on a Q-Exactive Plus Mass Spectrometer (Thermo Scientific).

MS1 spectra (30,000 resolution) were collected in the Orbitrap and data dependent acquisition of MS/MS spectra were obtained in the LTQ by collision induced dissociation (CID). The SageN Sorcerer SEQUEST 3.5 algorithm was used to search and match MS/MS spectra to a complete semi-tryptic human proteome database (REFSEQ (Human, version 62)) including pseudo-reversed decoy sequences (Elias and Gygi 2007, Xu, Duong et al. 2009). Searching parameters included precursor ion mass tolerance ( $\pm 20$  ppm), partial tryptic restriction, fixed mass shift for modification

**Gel #1****Gel #2**

**Figure 2.4. Coomassie stained SDS gel of samples for insoluble proteome analysis described in chapter 4.** Alkylated detergent-insoluble samples were separated on a 10% SDS gel and stained with Coomassie Blue G-250. Each sample lane was cut into five gel bands (represented by letters A-E above) corresponding to molecular weight ranges. Gel pieces were then digested overnight with trypsin at 37C for peptide generation for LC-MS/MS analysis. Aside: Control (1-6) and ALS (7-12) samples were run on Gel 1, while FTDALS (1-6) and FTD (7-12) samples were run on gel 2. First lane (unmarked) is MW ladder.

of carbamidomethylated Cys (+57.0215 Da) and dynamic mass shift for oxidized Met (+15.9949) that accounts for methionine oxidation modification. Only *b* and *y* ions were considered during the database match. There is a bias for *y* ions because these include charged arginine/lysine on the c-terminal end. In addition,  $X_{\text{corr}}$  and  $\Delta Cn$  were dynamically increased for groups of peptides organized by a combination of trypticity (fully or partial) and precursor ion charge state to remove false positive hits and decoys until achieving a false discovery rate (FDR) of < 1%.  $Cn$  is the normalized correlation score; it is a measure of how well the theoretical *y/b* ion spectrum correlates to the experimental spectrum and is corrected for peptide length (McHugh and Arthur 2008).  $X_{\text{corr}}$ , the cross correlation, represents the uncorrected value.  $Cn$  ranges between 1 and 0 (normalized), with 1 being perfect correlation between experimental and theoretical spectra and a score of 0 indicative that there are no aligned peaks between the spectra.  $\Delta Cn$  is the difference between the top-ranking  $Cn$  and the  $Cn$  for each other peptide; it is a score that gives a measure of how far the top score is above the other candidates (McHugh and Arthur 2008). Protein quantification was performed based on the extracted ion current (XIC) measurements of identified peptides as previously reported and by spectral counting (Donovan, Higginbotham et al. 2012, Dammer, Duong et al. 2013). Ion intensities for identified peptides were extracted in full-MS survey scans of high-resolution and a ratio of the peak intensities for the peptide precursor ion was calculated using in-house software as previously published (Gozal, Duong et al. 2009, Seyfried, Gozal et al. 2012, Dammer, Duong et al. 2013, Donovan, Dammer et al. 2013, Dammer, Lee et al. 2015). Accurate peptide mass and retention time (RT) was used to derive signal intensity for every peptide across LC-MS/MS runs for each case. The spectral count of a given protein is the summed number of its matching MS/MS spectra (Figure 2.5) (Lundgren, Hwang et al. 2010). Peptide ions are chosen for MS/MS analysis according to the



**Figure 2.5. Diagram of spectral count representation of relative protein abundance in LC-MS/MS.** The spectral count is a reasonable index that reflects protein abundance. The spectral count of a given protein is the summed number of its matching MS/MS spectra. As shown in the diagram, the relative abundance of the proteins in the sample, represented by different colors and shapes is determined after protein digestion and peptide analysis where the numbers of resultant spectra reflect the initial protein abundance in the sample.

rank of the ion intensity, so the spectral count of an identified protein reflects its abundance in the sample.

### *2.9 Total proteome sample preparation: homogenization and proteolytic digestion*

Dorsolateral prefrontal cortex (Brodmann area 9) tissue samples were processed essentially as previously described (Seyfried, Dammer et al. 2017). In brief, each piece of tissue was individually weighed (~100mg) and homogenized in 500ul of urea lysis buffer (8M urea, 100mM NaHPO<sub>4</sub>, PH 8.5), including 5ul (100X stock) HALT protease and phosphatase inhibitor cocktail (Thermo Fisher, Catalogue #78440). Homogenization was performed using a Bullet Blender (Next Advance) following manufacturer protocols. Each tissue piece was added to the urea lysis buffer in a 1.5mL Rino tube (Next Advance) that contained 750 mg stainless steel beads (0.9-2mm in diameter) and blended twice for 5 minute intervals in a 4°C cold room. Protein supernatants were transferred into 1.5mL Eppendorf tubes and sonicated (Sonic Dismembrator, Fisher Scientific) for 3 cycles of 5 seconds (at 30% amplitude) with 15 second intervals of rest to shear DNA. Samples were then centrifuged for 2 minutes at 15,871 G at 4°C. Protein concentration was assessed using the bicinchoninic acid (BCA) method, and samples were frozen at -80°C until use. Each homogenate was analyzed by SDS-PAGE to assess protein integrity (**Figure 2.6A**). One sample was run in replicate to show the consistency of the peptide quantification (**Figure 2.6B**). Brain protein homogenates (100µg) were diluted with 50mM NH<sub>4</sub>HCO<sub>3</sub> to a final concentration of less than 2M urea. Samples were subsequently treated with 1mM dithiothreitol (DTT) at 25°C for 30 minutes, and then 5mM iodoacetamide (Engelhart, Geerlings et al.) at 25°C for 30 minutes in the dark. Protein samples were digested with 1:100 (w:w) lysyl endopeptidase (Wako) at 25°C for 2 hours and then further digested with 1:50 (w/w)

trypsin (Promega) overnight at 25°C. Resulting peptides were desalted with a Sep-Pak C18 column (Waters) and dried under vacuum.

### *2.10 Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) and label free quantification*

Analysis by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) was performed using material and expertise available at the Emory University Proteomics core (<http://proteox.genetics.emory.edu/emory/core.html>). For the project described in chapter 4 the Q-Exactive(QE) Plus Orbitrap was used. The QE is coupled to nanoAcquity UPLC system. This instrument features high ion currents because of an S-lens, and fast high-energy collision-induced dissociation (HCD) peptide fragmentation because of parallel filling and detection modes. The e-image current from the detector is processed by an enhanced Fourier Transformation algorithm, enabling high mass spectrometric resolution (up to 140,000 fwhm). With almost instantaneous isolation and fragmentation, the instrument can sequence 10 peptides/sec. This instrument is also capable of a proteomics approach termed parallel reaction monitoring (PRM), which is a targeted proteomics strategy wherein all product ions of a peptide are simultaneously co-detected under conditions that offer high resolution and high mass accuracy. For the project described in chapter 5 the fusion Orbitrap hybrid mass spectrometer was used. This mass spectrometer is coupled to an UltiMate 3000 RSLCnano system. The Fusion is equipped with a mass filter, a collision cell, a high-field Orbitrap analyzer, and, finally, a dual cell linear ion trap analyzer (Q-OT-qIT). This system offers high MS/MS acquisition speed of 20 Hz and detects up to 19 peptides sequences within a single second of operation. The Fusion also has resolution in excess of 450,000 allowing for separation of isobaric interferences.

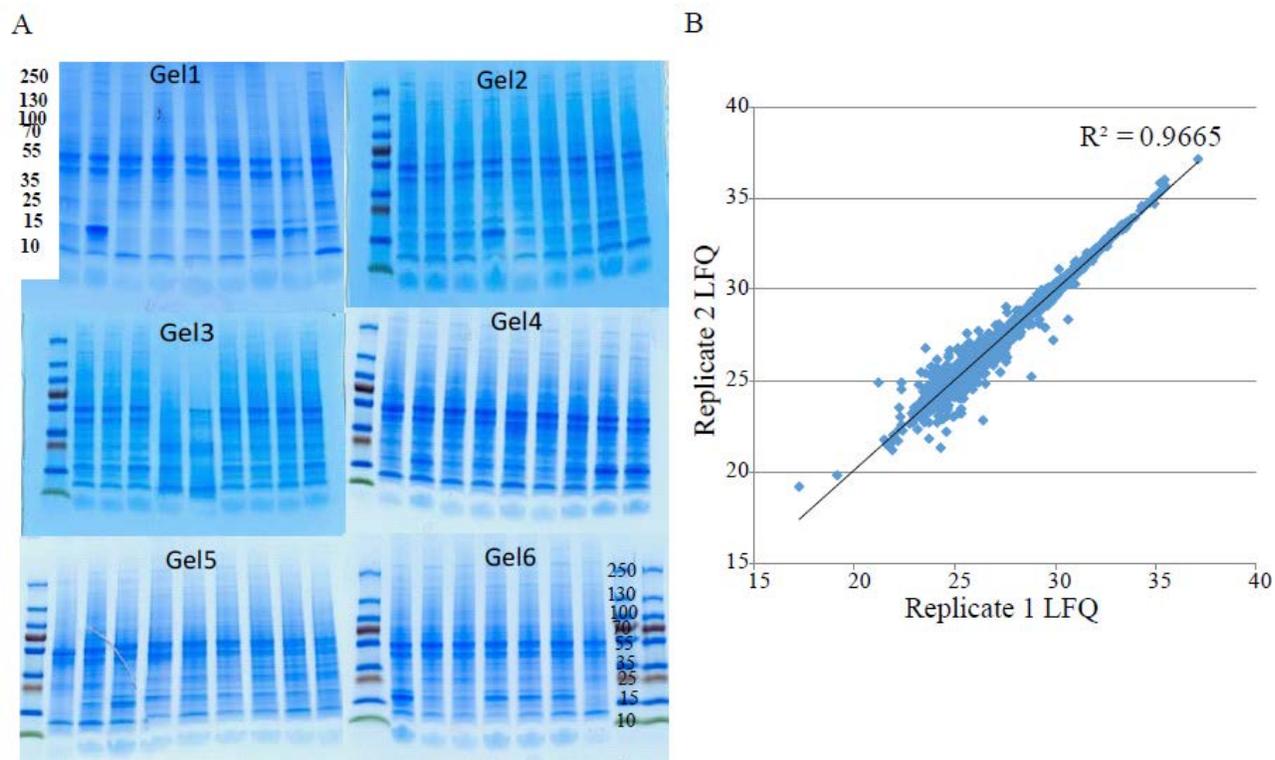
For LC-MS/MS analysis for total proteome analysis described in chapter 5, the peptides were first resuspended in 100  $\mu$ L of loading buffer (0.1% formic acid, 0.03% trifluoroacetic acid, 1% acetonitrile). Peptide mixtures (2  $\mu$ L) were separated on a self-packed C18 (1.9  $\mu$ m Dr. Maisch, Germany) fused silica column (25 cm x 75  $\mu$ M internal diameter (ID); New Objective, Woburn, MA) by a Dionex Ultimate 3000 RSLCNano and monitored on a Fusion mass spectrometer (ThermoFisher Scientific, San Jose, CA). Elution was performed over a 140-minute total gradient at a rate of 300 nL/min with buffer B ranging from 1% to 65% (buffer A: 0.1% formic acid in water, buffer B: 0.1 % formic in acetonitrile). The mass spectrometer cycle was programmed to collect at the top speed for 3 second cycles. The MS scans (400-1600 m/z range, 200,000 AGC, 50 ms maximum ion time) were collected at a resolution of 120,000 at m/z 200 in profile mode and the HCD MS/MS spectra (0.7 m/z isolation width, 30% collision energy, 10,000 AGC target, 35 ms maximum ion time) were detected in the ion trap. Dynamic exclusion was set to exclude previous sequenced precursor ions for 20 seconds within a  $\pm$ 10 ppm window. Precursor ions with +1, and +8 or higher charge states were excluded from sequencing.

For quantification, raw data files were analyzed using MaxQuant v1.5.2.8 with Thermo Foundation 2.0 for RAW file reading capability essentially as described with slight modifications (Seyfried, Dammer et al. 2017). The search engine Andromeda was used to build and search a concatenated target-decoy Uniprot human reference database. Protein Methionine oxidation (+15.9949 Da) and protein N-terminal acetylation (+42.0106 Da) were variable modifications (up to 5 allowed per peptide); cysteine was assigned a fixed carbamidomethyl modification (+57.0215 Da). Only fully tryptic peptides were considered with up to 2 miscleavages in the database search. A precursor mass tolerance of  $\pm$ 10 ppm was applied prior to mass accuracy

calibration and  $\pm 4.5$  ppm after internal MaxQuant calibration. Other search settings included a maximum peptide mass of 6,000 Da, a minimum peptide length of 6 residues, 0.6 Da Tolerance for ion trap HCD MS/MS scans. The false discovery rate (FDR) for peptide spectral matches, proteins, and site decoy fraction were all set to 1%. The label-free quantitation (LFQ) algorithm in MaxQuant (Luber, Cox et al. 2010, Cox, Hein et al. 2014) was used for protein quantitation as previously described. To account for possible confounds in run time, a brain peptide standard, generated from pooled samples of homogenized brain, was included at different points in the run to control for drift over time and highlight consistency in the protein measurements.

### *2.11 Statistics for clinical analyses*

Statistical analysis was carried out in IBM SPSS Statistics 22 (Chicago, IL). Between-group comparisons of continuous variables (age at onset, age at diagnosis, and time to diagnosis) were performed using Student t tests, and comparisons of categorical variables (sex, race, site of onset, presence of FTD, and positive family history of ALS) were analyzed by  $\chi^2$  tests. Kaplan- Meier analysis was used to determine effect of C9orf72 repeat expansion on survival from age at onset of symptoms to death (or tracheostomy). Significance was set at  $p = 0.05$  (2-tailed test). Multiple comparisons were accounted for by Bonferroni correction ( $p$  threshold after correction = 0.017). Other factors that may influence survival were adjusted for through a Cox proportional hazards model (backward stepwise likelihood ratio), including age at onset (as a continuous variable), sex (male vs female), site of onset (bulbar vs not bulbar), family history of ALS (familial vs sporadic), presence of FTD, and C9orf72 expansion status. Missing data in the analysis comparing presence of clinical FTD between the groups were handled by excluding cases from analysis as this was only assessed in 727 of the 781 cases.



**Figure 2.6. SDS-PAGE gel for total proteome analysis described in chapter 5.**

A. Each frontal cortex total homogenate was analyzed by coomassie-stained SDS-PAGE to assess protein integrity. Samples showed very little difference in proteome integrity, assessed by the banding pattern. B. A technical replicate was included in the LC-MS/MS run. Two Frontal cortex samples from the same case were homogenized and run separately. The LFQ abundance for these replicate samples show near perfect correlation ( $R^2=0.97$ ).

### *2.12 Statistics for pathological analyses*

Semi-quantitative neuropathological scores were compared between groups using Fisher's exact test to assess the neuropathological differences given our relatively small sample size. A p-value of  $\leq 0.05$  was defined as statistically significant. JMP Pro 13 statistical software was used for this analysis (SAS institute, Cary, NC, USA).

### *2.13 Statistics for analysis of insoluble proteome*

Unsupervised multidimensional scaling (MDS) analysis was conducted in R using an unsupervised clustering algorithm using Euclidean distance, Kendall Tau's rank correlation coefficient (power 9). MDS is a class of analysis, while principal component analysis (PCA) is one such method. The X and Y axes on the MDS plot represent the first and second principal components of expression data.

Differentially enriched or depleted proteins were identified by calculating the fold change difference ( $\pm 1.5$ -fold) and p-values from unpaired student's t-test. When multiple comparisons were performed a Bonferroni correction was done changing the significance threshold to account for the increased probability of spurious positives associated with multiple comparisons. Differential expression is sometimes presented as volcano plots, which were generated using Perseus, a software platform freely available through the Max Planck Institute of Biochemistry (Tyanova, Temu et al. 2016).

A netscreen was conducted to assess proteomic changes across the ALS-FTD spectrum that correlate with TDP-43 in the insoluble proteome. This is done by determining the bi-weight mid-

correlation (bicor) coefficient for all pairwise protein comparisons with the TARDBP LFQ from the insoluble proteome, respectively, across all 24 individual cases. Bicor is a median-based correlation measure and has been shown to be less susceptible to outliers than Pearson correlation (Langfelder and Horvath 2012).

## 2.14 Statistics for analysis of total proteome

### *Protein filtering and data imputation*

Protein abundance was determined by peptide ion-intensity measurements across LC-MS runs using the label-free quantification (LFQ) algorithm in MaxQuant (Cox, Hein et al. 2014). In total, 47,977 peptides mapping to 4,178 protein groups were identified. However, one limitation of data dependent label free quantitative proteomics is missing quantitative measures, especially for low abundance proteins (Karpievitch, Dabney et al. 2012, Seyfried, Dammer et al. 2017). Thus, only those proteins quantified in at least 90% of samples were included in the data analysis. After filtering, only allowing 10 percent (6) missing values maximum across the 51 LC-MS/MS runs, 2,612 unique proteins were identified and robustly quantified (**Figure 2.7B**). The 10% or fewer missing protein LFQ values were imputed using the k-nearest neighbor imputation function in R `impute::impute.knn()` function, similarly as previously described (Seyfried, Dammer et al. 2017).

### *Outlier removal and regression*

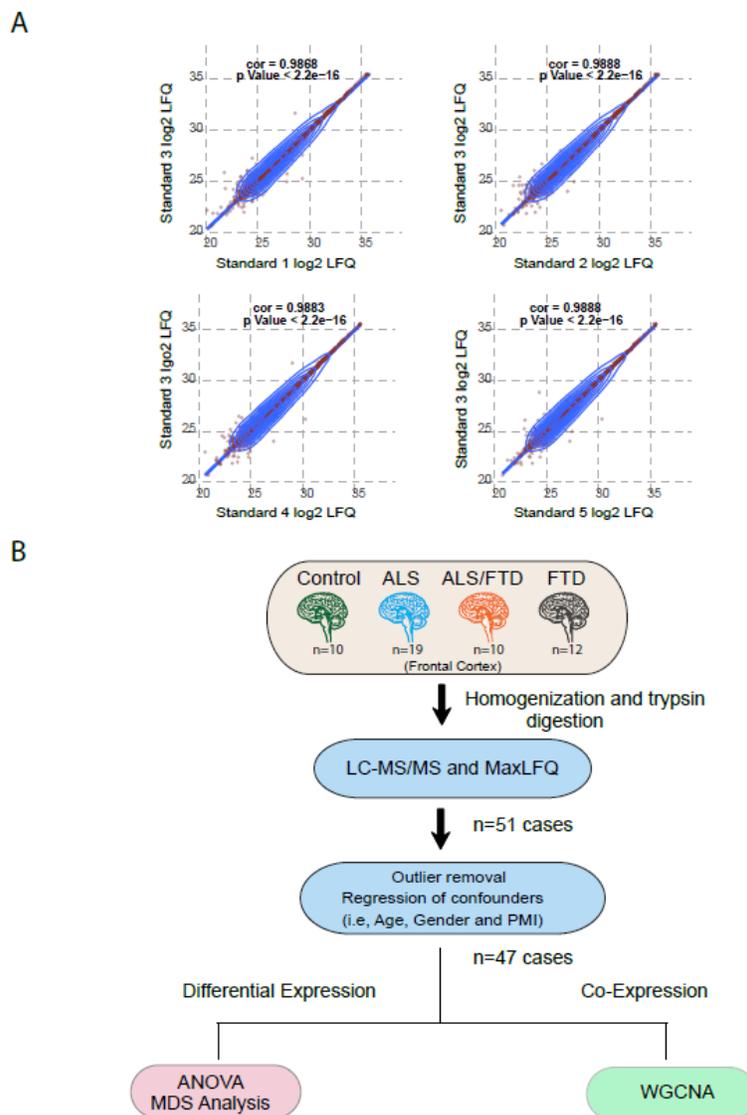
Prior to data analysis, outlier removal was performed using Oldham's 'SampleNetworks' v1.06 R script (Oldham, Konopka et al. 2008) as previously published (Seyfried, Dammer et al. 2017). Two control and two FTD cases were removed from the 51 cases initially included (**Figure**

**2.7A).** Bootstrap regression of the remaining 47-case LFQ intensity matrix that explicitly modeled case status category while removing covariation with age at death, gender, and postmortem interval (PMI) was done following principal component analysis(PCA) of the expression data to confirm appropriate regression of selected traits, both in the ‘SampleNetworks’ graphical output and via an in-house R script for PCA Spearman correlation to the amassed traits for 47 all non-outlier cases. PCA visualized that the top five principal components had Spearman correlation rho less than 0.3 with any of these three regressed covariates.

#### *Differential expression analysis*

Differentially enriched or depleted proteins ( $p \leq 0.05$ ) were identified by ANOVA comparing the four clinical groups (control, ALS, ALS/FTD and FTD). Multidimensional scaling as implemented in the WGCNA R package (Langfelder and Horvath 2008) was used to visualize separation of cases using a subset of 165 proteins which had at least 2 ANOVA Tukey pairwise comparisons of high significance ( $p < 0.01$ ) among the 6 possible comparisons among ALS, FTD, FTD/ALS, and control groups. In a separate analysis, differentially expressed proteins in C9orf72 expansion positive (C9Pos) ALS versus C9orf72 expansion negative (C9Neg) ALS (excluding cases with co-existing dementia) were identified by t-test. Differential expression is presented as volcano plots, which were generated with the ggplot2 package in Microsoft R open

3.3.2. Significantly altered proteins along with corresponding p-value are listed in **Table 5.1**.



**Figure 2.7. Proteomic quality control measures and data analysis pipeline**

**A.** Pooled homogenates of several neurodegenerative disease brains were used as a technical standard and analyzed at several points within LC-MS/MS batch. Correlation plots of LFQ data for each standard to the previous one show strong correlation. **B.** Workflow outline for data collection, and pre-processing prior to differential and co-expression analysis.

### *Co-expression network analysis*

Following previously described procedures of WGCNA (Seyfried, Dammer et al. 2017), a weighted protein co-expression network was generated using this pre-processed protein abundance matrix, using the `WGCNA::blockwiseModules()` function with the following settings: soft threshold power  $\beta=4.5$ ,  $\text{deepsplit}=4$ , minimum module size of 12, merge cut height of 0.07, signed network with partitioning about medoids (PAM) respecting the dendrogram and a reassignment threshold of  $p<0.05$ . Specifically, we calculated pair-wise biweight mid-correlations (bicor, a robust correlation metric) between each protein pair and transformed that matrix into a signed adjacency matrix (Langfelder and Horvath 2012). The connection strength of components within this matrix was used to calculate a topological overlap matrix, which represents measurements of protein expression pattern similarity across the set of samples in the cohort constructed on the pairwise correlations for all proteins within the network (Yip and Horvath 2007). Hierarchical protein correlation clustering analysis by this approach was conducted using 1-TOM, and initial module identifications were established using dynamic tree cutting as implemented in the `WGCNA::blockwiseModules()` function (Langfelder, Zhang et al. 2008). Module eigenproteins were defined, which represent the most representative abundance value for a module and which explain co-variance of all proteins within a module (Miller, Woltjer et al. 2013). Pearson correlations between each protein and each module eigenprotein were done; this module membership measure is defined as  $k_{ME}$ . **Figure 2.7B** illustrates workflow for analysis.

### *Module preservation and over-representation analyses*

Module preservation was tested using the ‘modulePreservation’ WGCNA R package function, using exactly 500 permutations comparing the frontal cortex proteomic network generated from samples in this study against a previously generated Emory human brain protein network, a similar network built with ALS and other neurodegenerative disease cases from the Emory brain bank (Seyfried, Dammer et al. 2017). This analysis was to ensure that the modules were representative of frontal cortex specific networks. Additionally, over-representation analysis (ORA) was conducted using Fisher Exact tests (two-tailed) between module membership of the ALS-FTD proteomic network versus the previous Emory brain protein network (Seyfried, Dammer et al. 2017). ORA analysis uses gene set enrichment analysis with a two-sided Fisher exact test with 95% confidence intervals by employing the R function ‘fisher.test’ (Seyfried, Dammer et al. 2017). To reduce false positives, FDR adjustment of p-values corrected for multiple comparisons.

#### *Enrichment analyses*

To characterize differentially expressed proteins and co-expressed proteins based on gene ontology annotation, we used GO Elite v1.2.5 as previously published (Seyfried, Dammer et al. 2017), with output visualized using an in-house R script (**Table 5.2**). Cell type enrichment was also investigated as previously published (Seyfried, Dammer et al. 2017). Enrichment of TDP-43 protein-protein interactions (Conforti, Barone et al.) across co-expression modules was investigated by intersecting module proteins with lists of genes known to interact with TDP-43, and assessing significance of overlap using a one-tailed Fisher Exact hypergeometric overlap test. TDP-43 PPI lists from BioGRID (<https://thebiogrid.org/117003/summary/homo-sapiens/tardbp.html>) and a previously published global analysis of TDP-43 interacting proteins

(Freibaum, Chitta et al. 2010) were used. The total list of identified protein groups was used as the background and the PPIs lists (**Table 2.6**) were filtered for presence in the total proteins list prior to cross-referencing. After assessing significance of TDP-43 PPIs, p values were corrected for multiple comparisons by the Benjamini-Hochberg method.

**Table 2.6 Fisher's exact test for PPIs enrichment analysis (Supplemental file 1)**

## 2.14 Rationale

The methods used to address the questions posed by this thesis are appropriate because they improve our understanding of these co-occurring diseases. Clinical and demographic analysis is typically conducted on a specific population as we did in our clinical cohort using previously collected information about patients and their disease progression (Boeve, Boylan et al. 2012, Chio, Borghero et al. 2012, Ratti, Corrado et al. 2012). Analysis of post mortem samples (by proteomics or transcriptomics) to better understand complex neurological diseases is a tool that has been used to study other neurological conditions, most notably Alzheimer's disease and autism (Donovan, Higginbotham et al. 2012, Hales, Dammer et al. 2016, Wu, Parikshak et al. 2016, Seyfried, Dammer et al. 2017). Mass spectrometry based proteomics is a method that allows us to utilize human samples to better understand pathways implicated in these diseases. Enrichment of the detergent insoluble brain proteome is a valid method to address the role of pathological inclusions along the ALS-FTD disease spectrum, as has been applied in other studies (Gozal, Duong et al. 2009, Hales, Dammer et al. 2016). Co-expression analysis goes beyond the typical differential analysis of proteomic data and moves us into a systems level analysis that allows us to get a better sense of networks that are altered in disease.

The studies included in this dissertation are not without limitations. In the work described in chapter 3, the most significant limitation is that our analyses are retrospective in nature. As we used information previously collected from patients, we were not able to compare features that were not initially assessed. The challenges posed by this are evident in our lack of information on psychiatric disease features, which may be a more prominent feature in C9orf72 ALS patients. Since we did not have this information in the majority of the cases we could not

examine the relationship between this clinical feature and the C9orf72 mutation. Another example of this limitation was reflected in our inability to conduct follow-up telephone interviews with our C9Neg patient cohort; this would have allowed for a comparison of how our definition of sporadic vs. familial cases may have changed in that group as well. Another challenge of the clinical analysis presented in this work is the fact that we were looking at a clinical cohort. Although this offers a rich database and substantial follow up, it affects the generalizability of the results as a clinical cohort will likely differ from a community cohort. Using previously acquired information on these cases also limited our assessment of the diagnosis of comorbid FTD, which prior to certain consensus criteria was based on clinical judgment and not formal cognitive testing, likely resulting in an underestimation of comorbid FTD in each of the groups. Still, we were able to exploit the rich resources available with DNA samples and linked demographic and clinical data from these patients to add a substantial report comparing C9Pos ALS patients and C9Neg ALS patients side by side. A strength of this work is that all of the patients included in this study are from a single clinic and provider, such that comparisons between patient presentations and clinical observations are very consistent and any inherent bias is equally distributed between the two groups. This work has significant implications for screening patients and incorporation of patients into clinical trials.

In the proteomic studies described in chapters 4 and 5 we also had several limitations. The richness and practicality of using human brains to study these human diseases cannot be overstated. However, the starting material for these studies is post mortem brain, which inherently means that the data generated reflects end stage disease. What this means is that all of our conclusions and analysis are based on end stage measures of the brain. In some ways, this

analysis is similar to a police officer showing up after an accident occurs and trying to piece together the pieces of the puzzle to figure out what happened. Nevertheless, this is still a useful tool to explore some of the questions that interest us, but all outcomes must be based in the context of this constraint. Another challenge for these studies is the selection of control groups. Our control cases had no known neurological diseases, but we do not know if other non-neurological diseases may have affected our results.

Other limitations related to the proteomic analyses are inherent to the technical approach we used. Data from shotgun proteomics are affected by a variety of known and unknown systemic biases and a high proportion of missing values (missing identifications or abundance values). Systemic bias is inherent in MS-based data because of the complex nature of samples and technical processing of samples; these all contribute to a non-biological signal that can affect results (Karpievitch, Dabney et al. 2012). Thus, it is important to remove excess technical variability. Several methods to remove systemic biases were included in our analyses. Normalization is one such technique. Normalization involves a global adjustment which forces the distribution of the measured values to center around a constant (the geometric mean in our case, the mean, median, or other fixed value for others). This assumes that most peptide abundances do not change, so the distribution of intensities across different samples should be similar. Missing data, either at the level of peptide identification or abundance, is a major challenge with MS data; it can be due to many reasons. The peptide may truly be present at detectable levels but could be incorrectly identified or not detected, the peptide could be present but below the threshold of identification, or the peptide may truly not be present in the sample. To deal with the first situation, using observed values to impute missing values is acceptable

(Callister, Barry et al. 2006). Another method to remove systemic biases is regression. ANOVA and regression models can efficiently remove systemic biases when the sources are known; this was used for total proteome data analysis. Nonetheless, high-throughput mass spectrometry is an extremely useful technology for protein identification and quantitation because of its ability to rapidly provide identification and quantitation of thousands of peptides (Aebersold and Mann 2003).

The two factors that challenge comprehensive protein identification by shotgun MS are the complexity of the proteomes being analyzed and the high dynamic range at which proteins are expressed in samples (Gstaiger and Aebersold 2009). Instrumental limitations of the mass spectrometer on dynamic range cause redundant identification of highly abundant proteins within a sample and hinder our ability to identify protein species of low abundance within complex samples. However, advancements in instrument capacities and advancements in fractionation techniques have improved our abilities for comprehensive proteomic analysis (Gstaiger and Aebersold 2009). Analysis by quantitative mass spectrometry for the detection of protein abundance using label-free quantification techniques has the advantage of having a greater dynamic range and achieving higher proteome coverage than techniques such as proteomics involving isotope-labelling methods which are thought to have the tradeoff with increased accuracy (Bantscheff, Schirle et al. 2007).

Label-free quantitative mass spectrometry methods are based on spectral counting or on peptide precursor ion intensities that are acquired from the first mass spectrometer (MS1) of a tandem mass spectrometer. Spectral counting is a semi-quantitative process for analyzing shotgun MS

data at a moderate mass resolution (Lundgren, Hwang et al. 2010). The spectral count method is built on the assumption that the rate of a peptide precursor ion selection for fragmentation in a mass spectrometer is correlated to its abundance. This technique depends on the quality of the MS/MS peptide identification, and though it is suitable and reliable for large and abundant proteins, the number of peptides observed from small proteins and low abundance proteins is often insufficient to allow for accurate quantification using this technique. Label-free quantification based on accurate mass and time signature of a peptide involves the alignment of high-mass accuracy spectra. Here peptides are identified across different LC runs based on their individual properties (specific retention time and precise mass to charge ( $m/z$ ) values), this allows for quantification of all peptides detected within samples that are within the sensitivity range of the analyzer independent of MS/MS acquisition (Gstaiger and Aebersold 2009). With this method, we simultaneously record MS signal intensities and identify peptides using MS/MS.

The C9orf72 cohort included in the total brain proteome analysis detailed in chapter 5 was important in addressing the question of the same genetic mutation can manifest in a variety of clinical phenotypes along the ALS-FTD disease spectrum. Genetic mutations can disturb the abundance and structure of proteins. Further, genetic mutations can also play an indirect role with global, wide-reaching effects on the abundance and expression of other cellular proteins (Gstaiger and Aebersold 2009). Comparing the proteomic profiles from brains of patients along the ALS-FTD spectrum can provide valuable information that will help us to identify pathways and networks involved in disease. This tool provides us with information at the molecular phenotype level, information which is not accessible using other methods and serves as a prudent next step in the quest to better understand these overlapping neurodegenerative conditions.

### **Chapter 3: Comparative analysis of C9orf72 and sporadic disease in an ALS clinic population**

*This chapter contains portions of a manuscript from Neurology, Vol 10, Umoh et al., Comparative analysis of C9orf72 and sporadic disease in an ALS clinic population. p.1024-30. doi: 10.1212/WNL.0000000000003067. © 2016 American Academy of Neurology*

## Introduction

The C9orf72 hexanucleotide repeat expansion is the most common genetic mutation identified in patients with amyotrophic lateral sclerosis (ALS). The expansion mutation is reported to be present in 40-50% of patients with a family history of ALS, and in 5-10% of patients with sporadic ALS (Byrne, Elamin et al. 2012, Mahoney, Beck et al. 2012, Majounie, Renton et al. 2012). It remains unclear whether patients with the expansion (C9Pos) are phenotypically different from patients without the expansion (C9Neg).

Patients with C9Pos ALS are reported to have a higher prevalence of bulbar onset, earlier age at onset, reduced survival, and a higher incidence of co-morbid dementia when compared to patients with sporadic disease (Murray, DeJesus-Hernandez et al. 2011, Chio, Borghero et al. 2012, Stewart, Rutherford et al. 2012, Irwin, McMillan et al. 2013, Snowden, Harris et al. 2013). Other studies, however, describe C9Pos ALS patients as exhibiting clinical features similar to those with sporadic disease, only with a more rapidly progressive disease course (Cooper-Knock, Hewitt et al. 2012, Sabatelli, Conforti et al. 2012). **Table 3.1** summarizes previous reports comparing C9Pos to C9Neg cohorts. Additionally, the pathological expansion has been identified in 5-10% of ALS patients designated as sporadic, though these analyses have excluded family histories of non-motor manifestations that are seen in people with the C9 expansion, specifically frontotemporal dementia (FTD) (Majounie, Renton et al. 2012, Ratti, Corrado et al. 2012, Sabatelli, Conforti et al. 2012).

There are practical advantages in defining the clinical similarities and differences between C9Pos and C9Neg patients, including generalization of mechanistic findings from C9 models to C9Neg ALS, and whether ALS clinical trials should focus specifically on the C9Pos population.

Study citation	Characteristics	Study Results (C9Pos, C9Neg)	Comparison
Byrne 2012 [3]	N	21 C9Pos, 170 C9Neg	
	Gender (female)	11 (53%), 69(41%)	NS
	Site of onset (bulbar)	7(33%), 57(34%)	NS
	Median age at onset	57.5 (50.5–63.5), 61.9 (56.4-67.9)	<b>S</b>
	Median survival	20 months, 26 months	<b>S</b>
Irwin 2013 [8]	N	64 C9Pos, 79 C9Neg	
	Gender (female)	25, 36	NS
	Site of onset (bulbar)	N/A	N/A
	Mean age at onset (SD)	55.1 (1.7), 60.3 (1.9)	<b>S</b>
	Survival	2.6±0.3 years, 3.8±0.4 years	<b>S</b>
Snowden 2013 [6]	N	11 C9Pos, 49 C9Neg	
	Gender (female)	6, 15	N/A
	Site of onset (bulbar)	5, 21	<b>S</b>
	Mean age at onset (SD)	60 (7), 62 (9)	NS
	Survival	N/A	N/A
	Family History	90%, 47%	<b>S</b>
Sabatelli 2012 [9]	N	69 C9Pos, 1688 C9Neg	
	Gender (female)	34 (47.1%). 718 (42.5%)	NS
	Site of onset (bulbar)	23 (33.3%), 432 (25.9%)	NS
	Median age at onset	59.0 (50.0-65.6), 62.8 (54.0-69.7)	<b>S</b>
	Survival (years) (95% CI)	2.7 years (2.1 to 3.3), 3.6 (3.3 to 3.8)	<b>S</b>
Garcia- Redondo 2013 [21]	N	67 C9Pos, 869 C9Neg	
	Gender (female)	44.8%, 43.4%	NS
	Site of onset (bulbar)	35.9%, 31.8%	NS
		55.24 ± 9.33 (26–71), 57.82 ± 14.50	

	Median age at onset $\pm$ SD	(10–95)	S
	Median survival (Burberry, Suzuki et al.)	41 months (34–47), 53 months (45–60)	S
	Family History	62.7%, 13%	S
	Comorbid FTD	32.8%, 12.5%	S
Borghero 2014 [22]	N	51 C9Pos, 220 C9Neg	
	Gender (female)	20 (39.2 %), 85 (38.6 %)	NS
	Site of onset (bulbar)	18 (35.3%), 55 (25.0%)	NS
	Median age at onset (IQR)	62.3 (55.0–67.4), 63.2 (55.9–71.4)	S
	Median Survival (IQR)	2.7 years (1.9–3.8), N/A	S
	FALS	33 (64.7%), 26 (11.8%)	S
	FTD	16 (31.4%), 19 (8.6%)	S
Majounie 2012 [1]	N	465 C9Pos, 3983 C9Neg	
	Gender (female)	233 (50.1%), 1732 (43.5%)	S
	Site of onset (bulbar)	139 (33.1%), 933 (26.0%)	S
	Median age at onset (Range; SD)	56.8 (27.0–80.0; 9.1), 58.7 (4.0–93.0; 12.8)	N/A
	Median Survival (IQR)	N/A	N/A
	Family History	221 (47.5%), 367 (9.2%)	S

**Table 3.1. Tabular comparison of studies reporting on clinical and demographic characteristics in C9Pos patients compared to C9Neg patients.** Study designs and analyses varied across the different reports. Comparison category refers to whether a significant difference was noted (S=Significant, NS=not significant, NA=information not available).

The Emory ALS Center maintains a large collection of DNA samples from ALS patients along with their demographic and clinical characteristics (Traxinger, Kelly et al. 2013). In this large clinical cohort, we screened patients' DNA for the C9orf72 expansion and compared the baseline characteristics of C9Pos and C9Neg ALS patients as well as their survival profiles to determine if C9Pos ALS is phenotypically distinct from C9Neg ALS.

Moreover, few studies have investigated whether any neuropathological similarities or differences exist among expansion mutation carriers. A major pathological hallmark in many sporadic and some genetic forms of ALS and the major subtype of FTD cases is aggregation of the transactive response DNA binding protein 43 kDa (TDP-43). TDP-43 is a major protein found within ubiquitinated inclusions in both ALS and FTD patient post mortem brains (Arai, Hasegawa et al. 2006, Neumann, Sampathu et al. 2006). TDP-43 is an RNA and DNA binding protein that regulates transcription and splicing. In ALS and FTD, TDP-43 gets hyperphosphorylated, cleaved to generate C-terminal fragments and mislocalized from the nucleus to the cytoplasm (Lee, Lee et al. 2012, Janssens and Van Broeckhoven 2013). Interestingly, in most non-C9orf72 ALS and FTD cases, phosphorylated TDP (pTDP-43) is seen in nearly all ubiquitinated inclusions, but in C9orf72 related cases the presence of pTDP-43 was not as consistent (Cooper-Knock, Hewitt et al. 2012). Several labs have found that ALS and FTD C9orf72 expansion mutation carriers display abundant p62-positive, pTDP-43 negative neuronal intra-nuclear inclusions in hippocampus and cerebellum (Al-Sarraj, King et al. 2011, Hsiung, DeJesus-Hernandez et al. 2012, Simon-Sanchez, Dopper et al. 2012, Snowden, Rollinson et al. 2012, Stewart, Rutherford et al. 2012). P62 is a protein encoded by the SQSTM1 gene, it is an autophagosome cargo protein that acts as an autophagy receptor, it has been associated to a

multitude of other cellular functions, including cellular differentiation, the immune response, apoptosis and endosomal organization. P62 immuno-reactive inclusions were overrepresented in neuronal cytoplasmic inclusions and glial inclusions in non-motor regions in C9orf72 expansion mutation cases in a study that concluded that pathology in non-motor frontal cortex regions and in the hippocampal CA4 subfield discriminated expansion mutation cases strongly from the rest of the ALS cohort (Cooper-Knock, Hewitt et al. 2012). p62, sequestome-1, is a protein that transports cargo for degradation and protects certain proteins from degradation by sequestering poly-ubiquitin chains; it is involved in many cellular processes such as selective autophagy, the oxidative stress response and cell signaling (Bjørkøy, Lamark et al. 2009). A clearer picture of the pathology underlying expansion mutation cases will allow additional hypotheses of disease pathogenesis specific to these patients to be generated and tested.

To determine if neuropathological differences exist between expansion mutation carriers and other ALS and FTD subtypes a descriptive immunohistochemical (IHC) study was conducted using postmortem samples from patients with and without the C9orf72 expansion mutation. Patient samples were selected based on TDP-43 immunoreactivity; cases with robust TDP-43 and light TDP-43 pathology within the hippocampus (as noted by colleagues at the Emory brain bank) were analyzed. Selected cases were stained using for P62 using single label and double label immunofluorescence techniques, as well as 3,3'-diaminobenzoic acid. Abundance of p62 inclusions in hippocampal and cerebellar sections was assessed. Co-localization of p62 and pTDP-43 pathology was also evaluated. The goal of these studies was to better characterize and distinguish neuropathology within clinically characterized C9orf72 expansion mutation carrier post mortem samples available through the Emory brain bank.

Based on previous reports, we hypothesized that cases with the C9orf72 expansion mutation (C9Pos) will have p62 positive inclusions that are not immuno-reactive by pTDP-43, and these cases will have distinct p62 pathology pattern in certain analyzed regions. We hypothesized that co-localization of pTDP and p62 will be identified more frequently in patients without the expansion mutation. We also predict that our C9Pos ALS clinical cohort will have distinct clinical features and reduced survival compared to our C9Neg ALS cohort.

## Results

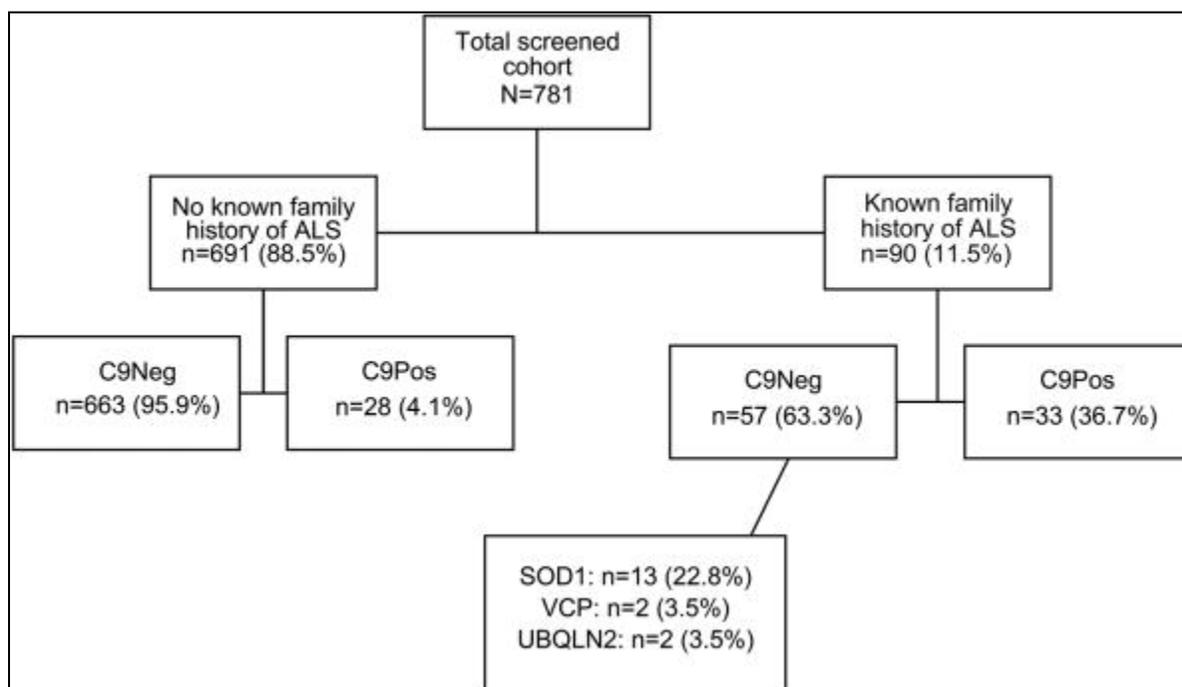
A total of 859 patients donated DNA for research, of which 781 samples were available for screening. All patients treated at the Emory ALS clinic between 2001 and 2015 were offered the opportunity to donate blood and clinical information for research purposes, half of the patients agreed and were included. The expansion mutation was identified in 61 cases (7.8%). Comparing the C9Pos and C9Neg populations, there were no differences in age at onset or diagnosis, race, and site of onset (bulbar vs. other) (**Table 3.2**). There was a shorter time to diagnosis in the C9Pos group (1 year vs. 1.5 years,  $p < 0.001$ ). There was also a difference in the sex ratio (M: F) of the two groups; 1.81 in the C9Neg group vs. 0.97 in the C9Pos group ( $p = 0.027$ ). Clinical symptoms consistent with FTD were identified in 21 patients among 727 ALS patients for whom cognitive or behavioral information was available. FTD was more common in C9Pos than C9Neg ALS patients (9/61 vs. 12/666,  $p < 0.001$ ). This was certainly expected given the understanding that the C9orf72 expansion is the most common mutation in patients with familial FTD (Hodges 2012).

A family history of ALS was originally recorded for 90 of the 781 patients examined (11.5%), 33 of these patients with a family history had the C9orf72 expansion (36.7%, **Figure 3.1**). Other disease-associated mutations accounted for an additional 19% of patients with familial disease, including 13 patients with SOD1, 2 patients with VCP, and 2 patients with UBQLN2 mutations.

	<b>C9Pos (n=61)</b>	<b>C9Neg (n=720)</b>	<b>P-value</b>
<b>Sex</b>	Male (30, 49.2%) Female (31, 50.8%)	Male (457, 63.5%) Female (263, 36.5%)	<b>0.027</b>
<b>Race</b>	Asian (0) Black (2, 3.3%) Caucasian (59, 96.7%) Other (0) Other-East India (0) Pacific Islander/ Native Hawaiian (0) Unknown (0)	Asian (3, 0.4%) Black (69, 9.6%) Caucasian (630, 87.5%) Other (5, 0.7%) Other-East India (1, 0.1%) Pacific Islander/ Native Hawaiian (4, 0.6%) Unknown (8, 1.1%)	0.572
<b>Family History of ALS</b>	True (33, 54.1%) False (28, 45.9%)	True (57, 7.9%) False (663, 92.1%)	<b>&lt;0.001</b>
<b>Site of Onset</b>	Bulbar (15, 24.6%) LE (19, 31.1%) UE (25, 41.0%) LE+UE (1, 1.6%) Diaphragm (0) Head Drop (0) Unknown (1, 1.6%)	Bulbar (191, 26.5%) LE (235, 32.6%) UE (279, 38.8%) LE+UE (5, 0.7%) Diaphragm (9, 1.3%) Head Drop (1, 0.1%) Unknown (0)	0.742

<b>Age at onset (years)</b>	Mean: 57.9 SD: 8.7	Mean: 56.3 SD:13.5	0.177
<b>Age at Diagnosis (years)</b>	Mean: 58.9 SD: 8.6	Mean: 57.8 SD:13.4	0.387
<b>Time to Diagnosis (years)</b>	Mean: 0.9 SD:0.8	Mean: 1.6 SD:2.1	<b>&lt;0.001</b>
<b>Survival (years)</b>  <b>Survival range (years)</b>	Median (95% CI): 2.4 (1.9-2.9) 0.1-7.0	Median (95% CI): 4.3 (3.8-4.7) 0.3-37.7	<b>&lt;0.001</b>
<b>Clinical FTD</b>	False (52, 85.2%) True (9, 14.8%)	False (654, 90.8%) True (12, 1.7%) Data not collected: (54, 7.5%)	<b>&lt;0.001</b>

**Table 3.2. Demographic and clinical characteristics comparing C9Pos ALS cases to C9Neg ALS cases.** P-values listed are from t-test, Pearson chi-square test, or Kaplan Meier survival log-rank analysis, see statistics section in methods. Chi-square analysis for site of onset compared bulbar onset to non-bulbar onset. SD: standard deviation. Survival is from age at symptom onset to death (or tracheostomy)

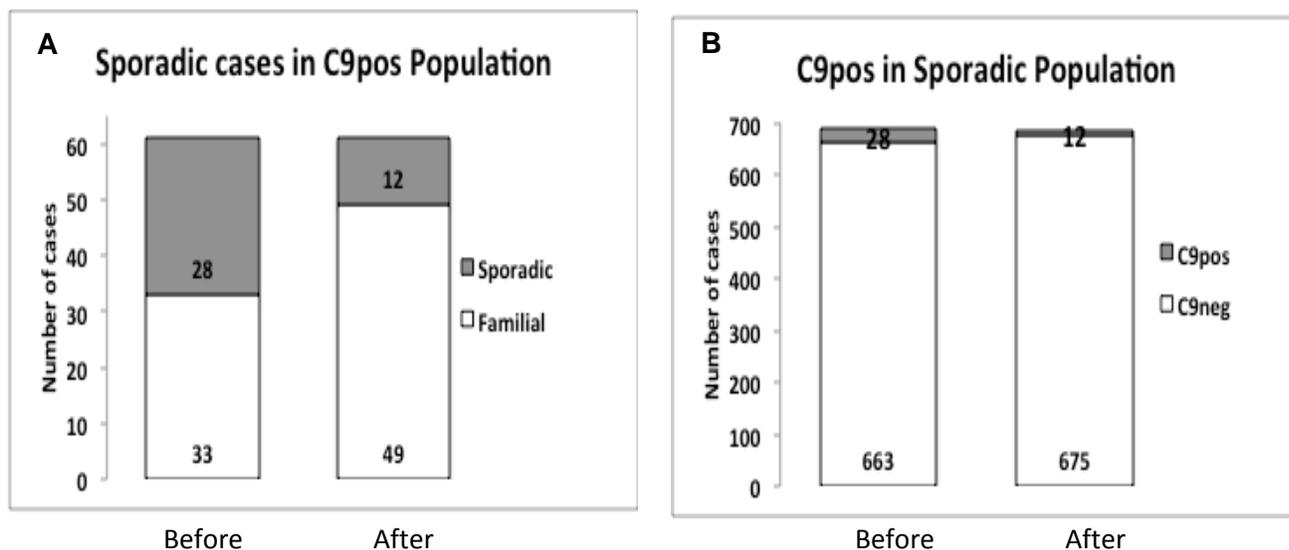


**Figure 3.1. Flowchart of family history of Amyotrophic lateral sclerosis (ALS) in screened cohort.** A total of 781 cases were screened for the C9orf72 expansion. Of those 781, 90 cases had a known family history of ALS, while 691 had no known family history of ALS. C9orf72 hexanucleotide repeat expansion (C9Pos) cases with a known family history of ALS made up 36.7% of the cases with a known family history of ALS. In the remaining cases with known family history of ALS who did not have the C9orf72 hexanucleotide repeat expansion (C9Neg), there were 13 patients with SOD1 mutations, 2 patients with VCP mutations, and 2 patients with UBQLN2 mutations. This chart represents information prior to follow-up interviews and reanalysis of clinical charts to obtain information on family history of dementia and other neurodegenerative diseases.

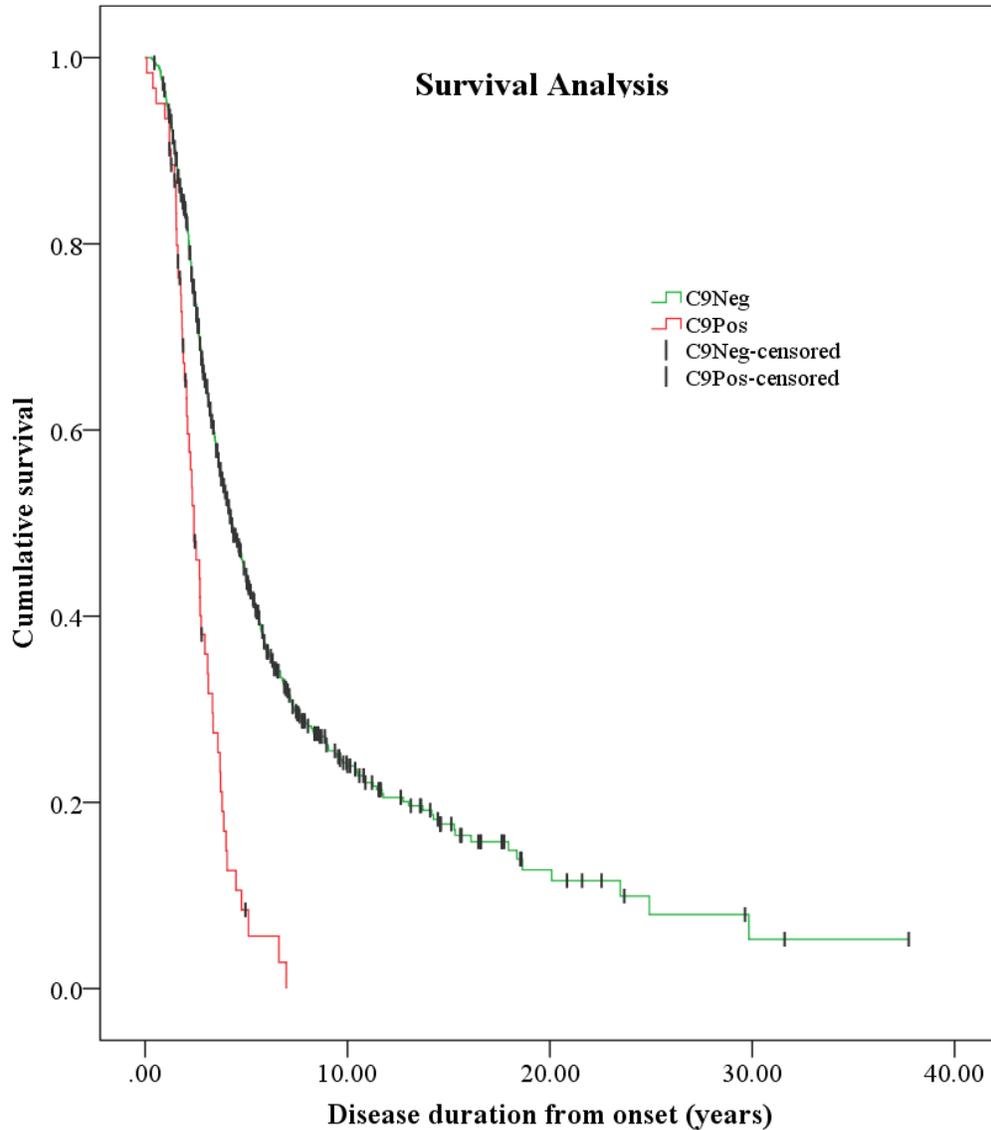
As expected, a family history of ALS was more common in C9Pos patients (33/61, 54%) than C9Neg patients (57/720, 7.9%) ( $\chi^2(1) = 117.63$ ,  $p < 0.001$ ). Twenty-eight of the 61 C9Pos patients had no family history of ALS (45.9%), indicating that 4.1% (28/691) of our sporadic population carried the C9orf72 expansion.

In contacting family members of the 28 C9Pos sporadic patients we identified one additional case with a brother who died with ALS (the patient did not previously provide this information). During this investigation, we also identified a family history of dementia in 16 of these 28 cases (57.1%), with 6 of these cases displaying behavioral changes consistent with FTD (Alberici, Cottini et al. 2012, Landqvist Waldo, Santillo et al. 2014). These changes include withdrawal, institutionalization, aggressive behavior, irrational behavior, and suicide attempts (**Figure 3.2**). Additionally, a family history of Parkinson's disease was identified in 2 of the 28 cases, one of which also had a family history of dementia.

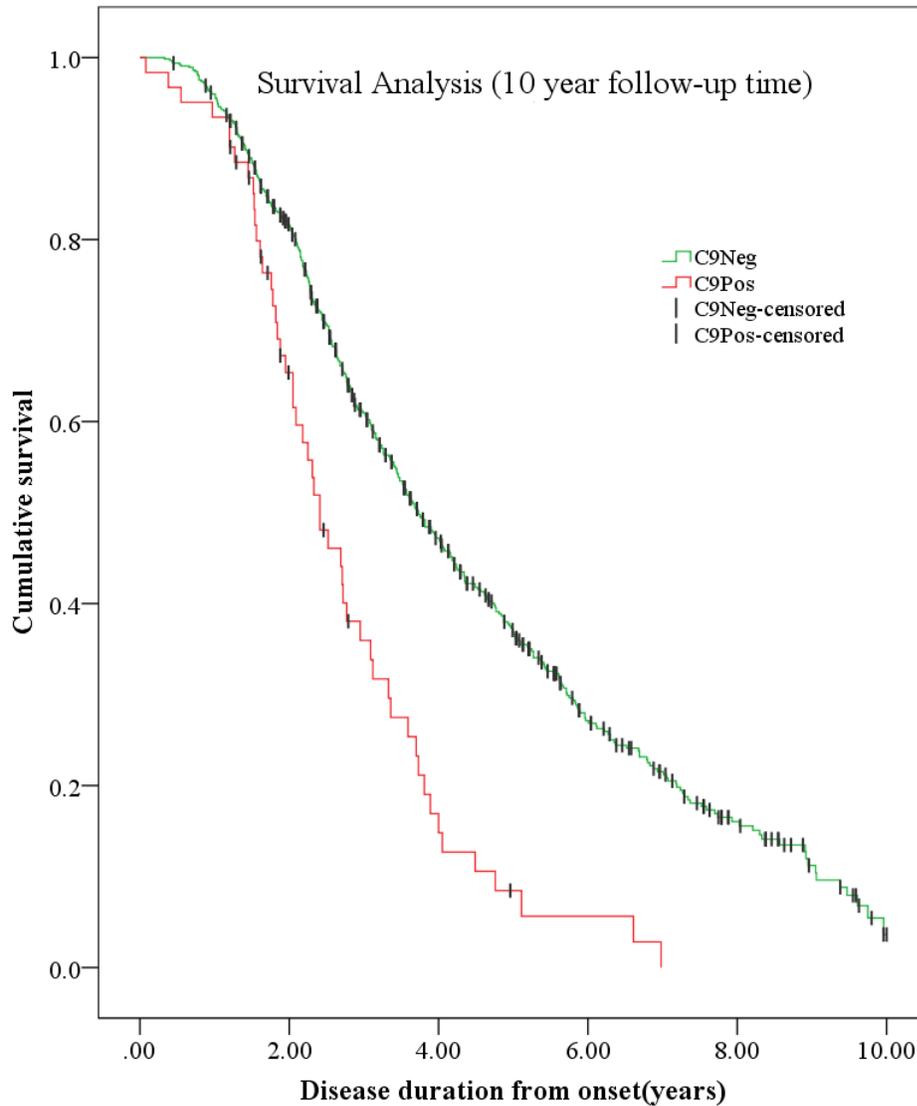
Kaplan Meier analysis showed a survival difference between C9Pos and C9Neg patients (log-rank  $\chi^2=45.323$ ,  $p<0.001$ ; **Figure 3.3**). Restricting the analysis to cases with survival or follow-up time less than or equal to 10 years, to more accurately represent ALS cohorts, revealed the same finding (**Figure 3.4**). A follow up Cox proportional hazards model that adjusted for age at onset, sex, site of onset, family history of ALS, FTD, and C9orf72 expansion status, confirmed that the survival difference was associated with the presence of the expansion mutation (HR:2.161, CI: 1.596-2.927,  $p<0.001$ ) (**Table 3.3**).



**Figure 3.2. Classification of familial and sporadic disease in the C9orf72 cohort.** A. Classification of family history in C9Pos cases prior to (before) and after reassessment. If we assume that a family history of dementia and other neurodegenerative disease suggests familial disease in a C9orf72 mutation carrier, then the number of sporadic cases is reduced from 28 (46%) to 12 (26%). B. The same reevaluation of family history applied to the sporadic population reduced the number of C9pos cases in this group from 28 (4%) to 12 (2%).



**Figure 3.3. Survival analysis comparing C9Pos and C9Neg ALS patients.** Kaplan-Meier curve analysis of survival of patients with amyotrophic lateral sclerosis with (C9Pos, red line) and without (C9Neg, green line) the pathogenic *C9orf72* expansion (log-rank [chi]  $2 = 45.323$ ,  $p < 0.001$ ). Hash marks indicate censored cases in each group. Overall, 781 cases (61 C9Pos, 720 C9Neg) were included in analysis. Median survival time for C9Pos = 2.4 years, 95% confidence interval (CI) 1.9-2.9, for C9Neg = 4.3 years, 95% CI 3.8-4.7.



**Figure 3.4. Survival analysis restricted to  $\leq 10$  years follow-up time.** Kaplan–Meier analysis of survival (from age at symptom onset) of patients with ALS with (C9Pos, red line) and without (C9Neg, green line) the pathogenic C9orf72 expansion, for the subset of patients who died within ten years of follow up (log-rank  $\lambda^2=30.207$ ,  $p<0.001$ ). Overall, 711 (61C9Pos, 650 C9Neg) cases were included in this analysis. Median survival time for C9Pos =2.4 years, 95% CI 1.9-2.9, for C9Neg=3.7 years, 95% CI 3.4-4.0.

Variable	Regression Coefficient(b)	Standard Error	p-value	e <sup>b</sup> Hazard ratio	95% CI for hazard ratio	
					Lower	Upper
Sex (0= female, 1= male)	0.075	.101	0.458	1.078	0.884	1.314
Site of onset (0= not bulbar, 1=bulbar)	0.506	0.102	<b>&lt;0.001</b>	1.659	1.358	2.027
Age at onset	0.030	0.004	<b>&lt;0.001</b>	1.030	1.023	1.038
Family History of ALS (0=no known family history of ALS, 1 = known family history of ALS)	0.161	0.149	0.283	0.852	0.635	1.142
Clinical FTD (0= false, 1= true)	0.679	0.249	<b>0.006</b>	1.971	1.209	3.214
C9orf72 expansion status (0=C9Neg, 1=C9Pos)	0.771	0.155	<b>&lt;0.001</b>	2.161	1.596	2.927

**Table 3.3. Cox regression model fitted to demographic and clinical data from C9Pos vs.**

**C9Neg patients.** Total number of cases included in analysis: 727 (complete information was available for 727 cases). Variables entered at step number 1: sex, site of onset bulbar, age at onset, family history of ALS, clinical FTD, and C9orf72 expansion status. Variable removed at step number 2: sex. Variable removed at step number 3: family history of ALS. Beginning at block number 1 method: backward stepwise (likelihood ratio).

Abbreviations: CI: confidence interval.

To understand whether this series is representative of the whole population attending the ALS clinic we analyzed characteristics of those cases that did not donate blood for research endeavors. For the patients seen at the clinic during the study interval that did not donate blood (i.e. those not included in the study), we pulled all available clinical data in order to compare this group to that reported in the manuscript. Because this group did not sign research consents, and thus did not fill out research questionnaires, we do not have as much detailed information on family histories. However, the groups were generally quite similar. Since these patients did not provide DNA or provide consent we did not include these data in the report.

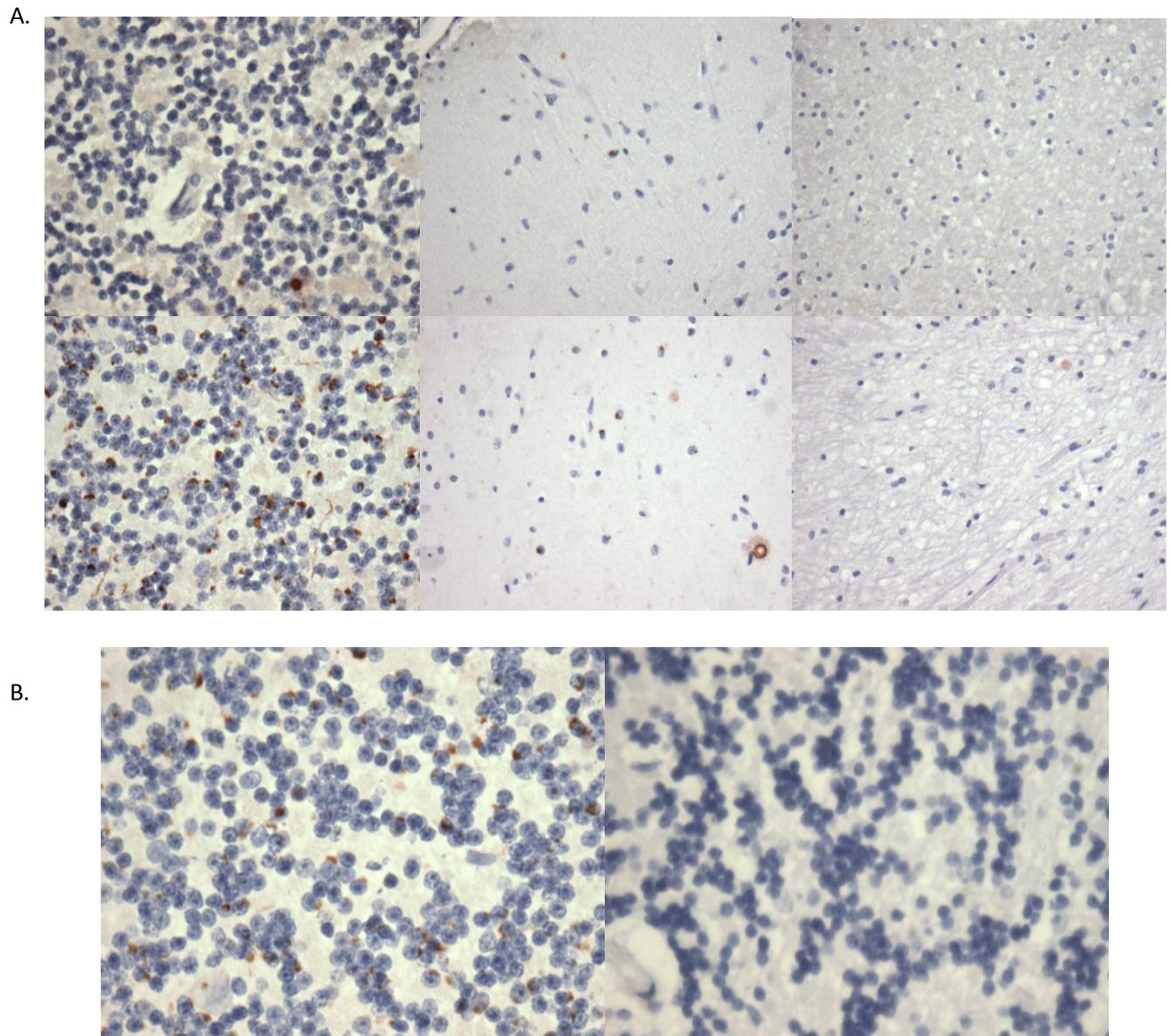
A subset of ALS cases (n= 20) were selected for double label immunohistochemistry using pTDP and p62 antibodies on cerebellar sections. Seven cases were found to have distinct cerebellar p62 pathology with abundant p62 staining in granule cell layers, mild p62 staining in molecular layers, and very little p62 immunoreactivity in the white matter or in Purkinje cells. All p62 immunoreactive inclusions within cerebellar sections were pTDP negative. After the study was unblinded it was shown that all 7 cases with distinct cerebellar p62 pathology were C9orf72 expansion mutation cases; demonstrating that p62 positive, pTDP negative cerebellar inclusions are pathognomonic for the C9orf72 cases. 8 cases with C9orf72 expansion mutation were included in this study; the one case that was not identified based on cerebellar pathology was noted by others at the brain bank to not stain with other common IHC stains most likely due to a different fixation technique immediately following tissue procurement. In essence, p62 positive pTDP negative cerebellar immunoreactivity gave a specificity of 100% and sensitivity of 87.5% for C9orf72 genotype. Interestingly, hippocampal sections showed p62 immunoreactivity in most cases regardless of genotype; but the C9orf72 expansion mutation

carriers tended to have many more p62 immunoreactive inclusions. Also, p62 inclusions in hippocampal sections of non-expansion mutation carriers was often localized in the region outside of the dentate gyrus while that of C9Pos cases were abundant and in high concentration within the dentate gyrus (**Figure 3.5**). Cases included in this analysis had different degrees of pTDP pathology within the hippocampus. Using double label IHC techniques levels of co-localization of both pTDP and p62 pathology were assessed. Cases were analyzed and categorized based on light pTDP-light P62, light pTDP- heavy P62, moderate pTDP-light P62, moderate pTDP-heavy P62, heavy pTDP-light P62, and heavy pTDP- heavy p62. These categories were defined by visually examining stained slides for amount of inclusion pathology. Co-localization of p62 and pTDP within hippocampal sections was seen in both C9orf72 carries and non-expansion mutation carriers. The abundance of p62 made it difficult to note any differences. Nevertheless, many of the C9Pos cases segregated into the light pTDP-heavy p62 or moderate pTDP-heavy p62 categories.

## **Discussion**

The recent discovery of the C9orf72 expansion mutation identified a distinct population of ALS patients that many have suggested are different, clinically and pathophysiologically, from other ALS populations (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011). This is an important question when designing and testing therapeutic interventions that may be specific for the C9Pos population. We approached this question by comparing baseline demographic and clinical features of C9Pos to C9Neg patients from a cohort of 781 ALS patients to determine whether these two groups are clinically distinct, adding to published reports from

European cohorts(Garcia-Redondo, Dols-Icardo et al. 2013, Borghero, Pugliatti et al. 2014). This was a clinic-based series



**Figure 3.5. Representative images of p62 immunohistochemical analysis of cerebellar sections.** A. p62 staining in two C9Pos cases throughout different cerebellar regions: granule cell layer (left), molecular layer (Nishimura, Mitne-Neto et al.), and white matter layer (right); 40X magnification. B. p62 Immunoreactivity within cerebellar sections is pathognomonic for C9orf72 Expansion. Cerebellar granule cell layer from C9Pos (left) and C9Neg (right) case; 40X magnification shown.

of ALS patients, which allowed us to avoid bias in patient selection. All patients at the Emory ALS center are asked to donate DNA for research purposes, and thus the only criteria for inclusion were the diagnosis of ALS and the consent to donate blood for research. Analysis of demographic and clinical characteristics from the cohort of patients seen at clinic that did not donate blood during the interval of this study showed that the patients included in this study are representative of the entire population attending the ALS clinic. Previous descriptions of C9Pos patients are relatively small and do not include comparisons to a large, unselected cohort of C9Neg patients (**Table 3.1**)(Byrne, Elamin et al. 2012, Montuschi, Iazzolino et al. 2015).

Other than the presence of a family history of ALS and increased prevalence of FTD, these two groups were clinically similar at presentation to the clinic. The distribution of site of onset of disease in our cohort of C9Pos patients closely mirrored that of C9Neg patients and was typical of other ALS cohorts (Williams, Fitzhenry et al. 2013). The racial makeup of both groups was largely Caucasian (Chio, Logroscino et al. 2013), which reflects our patient population. The equal distribution of men and women in the C9Pos group may reflect the autosomal dominant inheritance of the C9orf72 expansion mutation. Age at disease onset and age at disease diagnosis were similar and comparable to published reports from other ALS cohorts (Freer, Hylton et al. 2015, Valle, Roberts et al. 2015). Interestingly, C9Pos patients were more likely to receive a diagnosis sooner after symptom onset than C9Neg patients. This difference may be explained by a high index of suspicion in patients with a known family history of ALS. This hypothesis is supported by a shorter time to diagnosis in all patients with a family history of ALS (1.2 vs. 1.6 years).

We did not identify an earlier age at disease onset or increased prevalence of bulbar onset as was previously reported for other C9Pos ALS cohorts (Chio, Borghero et al. 2012, van Rheenen, van Blitterswijk et al. 2012, Irwin, McMillan et al. 2013, Cooper-Knock, Shaw et al. 2014). Earlier age at onset in C9Pos patients has been used as rationale to describe C9orf72-related ALS as a more aggressive form of disease. Our inability to confirm this finding might be explained by our study design, which compared the C9Pos group to a concurrent and unselected group of ALS patients, rather than a designated comparison cohort (Irwin, McMillan et al. 2013). Additionally, the published reports of increased prevalence of bulbar onset in C9Pos patients has been used to suggest a more severe disease since patients with bulbar disease are typically thought to carry a worse prognosis (Magnus, Beck et al. 2002, del Aguila, Longstreth et al. 2003). Bulbar onset was not overrepresented in our C9Pos cohort, though we acknowledge that the numbers of C9Pos compared to C9Neg patients were relatively small.

As expected, we found that C9Pos patients were more likely to report a family history of ALS, and there was an increased prevalence of clinical FTD in this group. Since formal cognitive testing was not conducted, the number of identified cases with comorbid FTD was low compared to reports of FTD in other ALS cohorts; our analysis of FTD within each of the groups only identified the most obvious FTD cases and likely underestimates the true presence of comorbid FTD in our population. Our finding of reduced overall survival in our C9Pos population compared to C9Neg patients is consistent with the published experience from other C9Pos cohorts (Byrne, Elamin et al. 2012, Chio, Borghero et al. 2012, Irwin, McMillan et al. 2013, Borghero, Pugliatti et al. 2014, Montuschi, Iazzolino et al. 2015). Additionally, we used a Cox

proportional hazards model to determine the influence of variables other than the C9 mutation, such as co-morbid FTD, that may affect survival in ALS (Magnus, Beck et al. 2002, del Aguila, Longstreth et al. 2003, Hu, Shelnutt et al. 2013). Survival differences between the C9Pos and C9Neg groups persisted, indicating that the C9orf72 expansion is independently associated with survival.

The link between ALS and dementia has been established for some time, with reports many decades ago of both sporadic and familial ALS patients with comorbid FTD (Ziegler 1930, Finlayson and Martin 1973). The discovery in recent years of genetic mutations underlying the coexistence of ALS and FTD in families, the most common of which is the C9orf72 repeat expansion, has advanced our knowledge of ALS as a multi system disease (Devenney, Vucic et al. 2015). Yet our clinical definitions for family history of disease have changed very little, and clearly should include questions about dementia, especially with behavioral changes, as well as other neurodegenerative phenotypes. Our re-analysis of family history in our C9Pos cohort with no known family history of ALS suggests that relatively few of these patients lacked any family history of neurodegenerative disease (**Figure 3.2**). Consideration of these new data reduced the prevalence of C9orf72 mutations in the sporadic ALS population to less than 2%, which is on par with reports of SOD1 mutations in the sporadic ALS population (Pasinelli and Brown 2006, Conforti, Barone et al. 2011). When SOD1 mutations were first identified as causative in ALS, a number of cases were considered to be sporadic (Jones, Brock et al. 1993). With further consideration of family histories, some of these cases were reclassified as familial, leading to our current understanding that an exceedingly small percentage of sporadic cases carry SOD1 mutations. Our results and knowledge of standard methods of defining familial disease cautions

us to be more thoughtful in classifying ALS as “sporadic” or “familial”, especially when using historical cohorts.

The C9orf72 expansion phenotype may be different even within families. In individual kindreds some relatives have dementia while others have ALS or both ALS and dementia (DeJesus-Hernandez, Mackenzie et al. 2011). Age at disease onset and survival within a single family with the C9 expansion is also very heterogeneous (Cooper-Knock, Shaw et al. 2014). It has been suggested that C9orf72 expansions are fully penetrant by age 80, though there are reports of C9Pos individuals over age 80 without a clinical phenotype (Majounie, Renton et al. 2012, Galimberti, Arosio et al. 2014). Within our C9Pos cohort age at onset of ALS ranged from 34 to 73 years, showing the wide heterogeneity even within this population. Also, though reduced survival is present in our C9Pos group compared to the C9Neg group, this may suggest that previous assumptions of expansion carriers being drastically different than other ALS patients are open to question.

A limitation of this study was our inability to conduct follow-up telephone interviews with our C9Neg patient cohort, as this would have allowed for a comparison of how our definition of sporadic vs. familial may have changed in that group as well. Also, the diagnosis of comorbid FTD was based on clinical judgment and not formal cognitive testing, likely resulting in an underestimation of comorbid FTD in each of the groups. Another limitation is the retrospective nature of our analysis. Due to this we were unable to assess the presence of specific psychiatric manifestations in our C9orf72 cohort. A strength of this work is that all of the patients included in this study are from a single clinic with one provider, such that comparisons between patient

presentations and clinical observations are very consistent and any inherent bias is equally distributed between the two groups.

Overall, the similarities between C9Pos and C9Neg patients are the most interesting outcome from this work. It begs us to ask why this disease clinically looks so similar in both groups, and whether common pathophysiology underlies disease expression in C9Pos and C9Neg patients. To elucidate common pathways implicated in ALS such that genetic and sporadic forms produce similar clinical phenotypes will require deeper genetic, molecular, and proteomic analyses. This report presents characteristics of C9Pos ALS patients and C9Neg ALS patients side by side and identifies remarkable demographic and clinical similarities that exist in the context of significant survival differences. These results are important for screening patients and incorporation into clinical trials.

Using our autopsy cohort, we confirmed that p62-positive, pTDP-negative immunoreactive cerebellar inclusions are common in C9orf72 associated ALS and FTD cases. We did not identify a distinct pattern of co-localization of p62 and pTDP in hippocampal sections as has previously been reported (Troakes, Maekawa et al. 2012). This study alludes to a larger issue in the field: pTDP-43 has been identified as the major pathological protein in both ALS and FTD, but there are patients who do not have pathological pTDP-43 inclusions and others have varying amounts of pTDP-43, like the C9orf72 cases. Neuropathological hallmarks of disease serve as clues of pathological mechanisms underlying disease. For instance, do our results mean that in patients with the expansion mutation the development of MND and dementia is different than

other ALS and FTD patients? Are factors that mediate the convergence of these diseases in C9orf72 carriers different than those in sporadic cases?

Another issue based on our neuropathological analysis highlight is that prior to the identification of the C9orf72 expansion, unique cerebellar neuropathology was not identified in ALS and FTD patient post mortem brain samples and the cerebellum was not considered to be implicated in these diseases. p62 cerebellar pathology seen in C9orf72 cases may clue us into protective mechanisms present in the cerebellum. Future studies investigating p62 interactions, with other proteins previously identified as pathological, will be useful in elucidating pathogenic mechanisms involved in disease development. Alternatively, cerebellar brain tissue sections were not typically analyzed in ALS and FTD cases so there may be pathology within this region that was previously overlooked. Neuropathological studies like this one are critical in elucidating pathways to neuronal dysfunction involved in ALS and FTD.

**Chapter 4: Alterations in the detergent-insoluble brain proteome linked to TDP-43 along the ALS-FTD disease spectrum**

## **Introduction**

Abnormal protein aggregation within the brain is a common pathological feature of many neurodegenerative diseases, including amyotrophic lateral sclerosis (ALS) and frontotemporal lobar dementia (FTD). The most common protein implicated in these disease is TAR DNA Binding protein (TDP-43), making many characterize these as diseases that lie along a spectrum termed TDP-43 proteinopathies. Previous studies relied on direct visualization of protein aggregates using standard immunohistochemistry protocols to identify proteins implicated in disease pathogenesis. However, present-day proteomic techniques enable us to utilize direct sequencing to identify proteins implicated in disease pathogenesis. Moreover, post mortem brain samples from individuals with these neurodegenerative phenotypes can undergo fractionation using a series of high speed ultracentrifugation steps to separate aggregated proteins within the insoluble fraction, these may represent proteins that are involved in disease either through an inability to maintain their normal function or a gain-of-function mechanism. Thus, further identification and delineation of the group of proteins enriched in the detergent insoluble brain fraction from cases along the TDP proteinopathy spectrum may give us important insights into disease pathogenesis.

Using post mortem brain samples to better understand neurodegenerative conditions used to be the strongest strategy of connecting diseases along the ALS-FTD continuum. Advances in genetics and deeper clinical phenotyping solidified the overlap of ALS and FTD. These diseases affect humans and are quite complex. As such, animal and cellular models of these diseases have many limitations (Burns, Li et al. 2015). Models that base disease pathogenesis on one aspect of the disease miss the complexity of these conditions. Using brain tissue from patients with

diseases along the ALS-FTD continuum and controls offers a chance to better understand the molecular underpinnings of this disease spectrum without oversimplifying them.

Proteomic sequencing of the detergent insoluble fraction from human postmortem brain has been performed by several groups to identify protein aggregates in other neurodegenerative diseases, mainly Alzheimer's disease. This endeavor has generated several targets and new insights into cellular pathways involved in Alzheimer's pathogenesis (Gozal, Duong et al. 2009, Hales, Dammer et al. 2016). Technological advancements in protein sequencing technology used for proteomic analyses has driven the field from the identification of a relatively small number of highly abundant proteins within samples that may have no disease relevance to a more sophisticated identification of proteins involved in disease pathogenesis (Gstaiger and Aebersold 2009). Since proteins can interact with all other classes of molecular components and can be modified in a variety of ways including their abundance, their pattern of post translational modifications and their proclivity to interact with other components in the cell, the global analysis of proteins in the brains of patients presenting with a certain phenotype is important (Gstaiger and Aebersold 2009). Also, since mRNA abundance has poor correlation with protein abundance, changes in protein abundances cannot just be inferred from DNA microarray data (Ideker, Thorsson et al. 2001).

Regarding TDP-43 proteinopathies, this technique that allows for the analyzation of proteins within samples could be very valuable. TDP-43 is a protein that lies at the center of ALS and FTD, two seemingly disparate diseases. ALS is a progressive neurodegenerative disease that involves the selective loss of motor neurons in the brain and spinal cord which leads to muscle

wasting, paralysis, and ultimately death within an average of 3 years of disease onset. The motor neuron loss is evident in the spinal cord and motor cortex of ALS patients. FTD on the other hand is a progressive neurodegenerative disease that involves the frontal and temporal lobes and is associated with changes in personality, language, behavior, and deficits in executive function.

The molecular mechanisms underlying TDP proteinopathies have mostly been identified from cellular and animal studies (Zhang, Xu et al. 2009, Lee, Lee et al. 2012, Janssens and Van Broeckhoven 2013). Aggregation of misfolded TDP-43 is a hallmark of ALS and half of FTD cases. TDP-43 is a highly conserved, nuclear RNA binding protein. TDP-43 is involved in transcription and regulation of splicing (Zhang, Xu et al. 2009). It is a protein that is normally concentrated in the nucleus but shuttles between the nucleus and cytoplasm. In disease, it accumulates in the cytoplasm. Further evidence supporting a direct causal link between ALS and TDP-43 brain proteinopathy was the identification of pathogenic mutations in the TARDBP gene in some ALS patients (Kabashi, Valdmanis et al. 2008, Sreedharan, Blair et al. 2008). These causal mutations in TARDBP were first identified in familial ALS cases and a few mutations have been described in ALSFTD and FTD patients. However, TDP pathology is not restricted to patients with genetic mutations in the gene. The function of TDP-43 in normal brain and neurodegenerative conditions is still elusive.

Previously, TDP-43, the protein encoded by TARDBP, has been shown to be involved in many steps of gene regulation, primarily with splicing and transport, but also in stability, translation, and microRNA maturation (Xu 2012). Mouse models expressing mutant TDP-43 have significant alterations in splicing and a motor neuron disease phenotype in the absence of

aggregation of TDP-43 or neuronal loss (Arnold, Ling et al. 2013). It is still debated whether TDP-43 acts through a gain of function or loss of function mechanism for its pathogenicity. TDP-43 binding partners have been assessed in cell culture models, mouse brains, and post mortem human brains from ALS and FTD patients. TDP-43 has been shown to preferentially bind long clusters of UG-rich sequences, mostly at intronic regions (Polymenidou, Lagier-Tourenne et al. 2011, Bhardwaj, Myers et al. 2013). Like other heterogeneous nuclear ribonucleoproteins (hnRNPs) TDP-43 autoregulates its expression levels by a negative-feedback loop. TDP-43 is a 414-amino acid protein encoded by the TARDBP gene on chromosome 1p36.2. It is highly conserved and ubiquitously expressed in many tissues including the brain (Zhang, Xu et al. 2009) (Mackenzie and Rademakers 2008). The structure of TDP-43 includes a nuclear localization signal (NLS), a nuclear export sequence (NES), two RNA-recognition motifs (RRM1 & RRM2), and a c-terminal glycine-rich region (GRR) (Zhang, Xu et al. 2009). Most of the mutations identified in TARDBP from patients suffering from familial or sporadic ALS are found in exon 6 which encodes the highly-conserved c-terminal domain of the gene, a region required for its self-regulation. Mutations in TARDBP are mostly predicted to enhance aggregation (Zhang, Xu et al. 2009). Several other RNA binding proteins have been associated with ALS and FTD, these include hnRNPA1, hnRNPA2/B1, hnRNPA3, TAF15, and EWSR1 (Donnelly, Grima et al. 2014). These associations are through the identification of pathogenic mutations in these genes in patients or findings that pinpoint them to pathological inclusions. Alteration in RNA-binding most likely plays a critical role in neurodegeneration in ALS and FTD patients.

One major problem in better understanding the role of TDP-43 in these diseases is the lack of pathology in animal models of disease. Numerous transgenic animal models expressing both wildtype and mutant TDP-43 have been created, and cytoplasmic inclusions were rarely observed though these animals develop an ALS-phenotype. Thus, although progress has been made in uncovering the biology of TDP-43 proteinopathy associated with neurodegeneration in ALS and FTD patients, the molecular and cellular mechanisms involved in the disease are still largely unknown. From genetic studies and pathological studies there is evidence to show that different disease pathways are involved; these include RNA processing, protein homeostasis, and TDP-autoregulation. Thus, it is likely that the disease pathomechanism involves many different pathways that converge to bring about the pathology.

We wanted to identify whether clinical phenotypes identified by clinicians were defined by different proteomic signatures, and what these signatures could inform us about disease pathogenesis. To do this, we used post mortem brain tissues derived from well-defined clinically characterized cases from the Emory University brain bank. The careful choice of cases with clinical information (**Table 2.3**) allowed us to determine whether pathological and proteomic differences determine clinical differences, something that has not previously been confirmed. We analyzed the detergent insoluble brain proteome (the frontal cortex) from ALS, ALSFTD, and FTD cases as compared to non-neurological disease controls to identify proteins that were enriched in TDP-43 proteinopathies and to identify proteins unique to each of these diseases. The purpose of this study was to identify and quantify proteins in the detergent-insoluble brain proteome correlated with TDP-43 pathology, as defined by label free quantification (LFQ) of TDP-43 using LC-MS/MS, across the clinical groups. To examine protein co-aggregation in

ALS and FTD linked to TDP-43 pathology, pairwise correlation coefficients were determined for each quantified protein to pTDP pathological rating in the frontal cortex, the C to N-terminal TDP-43 peptide ratio, and the LFQ of TARDBAP, across the 24 cases.

## Results

### **Label free proteomic analysis of detergent-insoluble proteome from individual cases along the ALS-FTD disease spectrum**

Samples of postmortem frontal cortex tissue sections from ALS, ALSFTD, FTD, and non-neurological disease control cases were collected and analyzed individually by label free proteomics (**Table 2.3**). Non-neurological disease control cases (n=6) were identified as clinically normal prior to death and did not have obvious cognitive decline. The average post mortem interval (PMI) of the groups ranged from 11 to 16 hours; there was no significant difference in this interval across the cases or controls. The age range of the individual cases was 48 to 94 years, and though the controls were older than the disease cases, in the disease groups there was no significant difference in age across the disease groups. Gender was equally balanced in each group (**Table 2.3**). There were no significant differences in disease duration across the clinical groups, though in our few samples ALS patients had the shortest disease duration, followed by ALSFTD patients. Samples were individually homogenized and enriched for sarkosyl-insoluble fraction before resolution by SDS-PAGE followed by coomassie staining (**Figure 2.4**). There were some differences in band intensities across the cases, but visual inspection showed that there was little degradation of the proteins which made major band patterns similar across the samples. An equal amount of protein from the sarkosyl-insoluble fractions was separated by SDS-PAGE, samples were then extracted from the gels and

subsequently digested with trypsin. Tryptic peptide mixtures were sequentially analyzed using identical LC-MS/MS conditions on a Q-Exactive Plus Mass Spectrometer (Thermo Scientific). The generated spectra were searched against a human protein database and filtered by mass accuracy and matching scores. Protein abundance was quantified by assessing spectral count data, the number of MS/MS events observed for a protein in the mass spectrometer, which has been shown to correlate strongly with the protein's abundance in a complex mixture (Lundgren, Hwang et al. 2010). Protein abundance was also quantified based on peptide-ion intensities. We sequenced 59,794 peptides to 7,293 proteins representing 6,915 unique protein groups. 1,130,970 peptide spectral counts were collected across all gel bands sequenced from the 24 individual cases. After filtering for a maximum of 10% missing data (due to shotgun approach) and normalization of spectral count data, we identified 3209 proteins (**Table 4.1**). Using extracted ion current data after filtering for missing data and normalization, we identified 4647 proteins. The relative abundance ( $\log_2$  transformed) for all detergent insoluble proteins, is shown in **Table 4.1**.

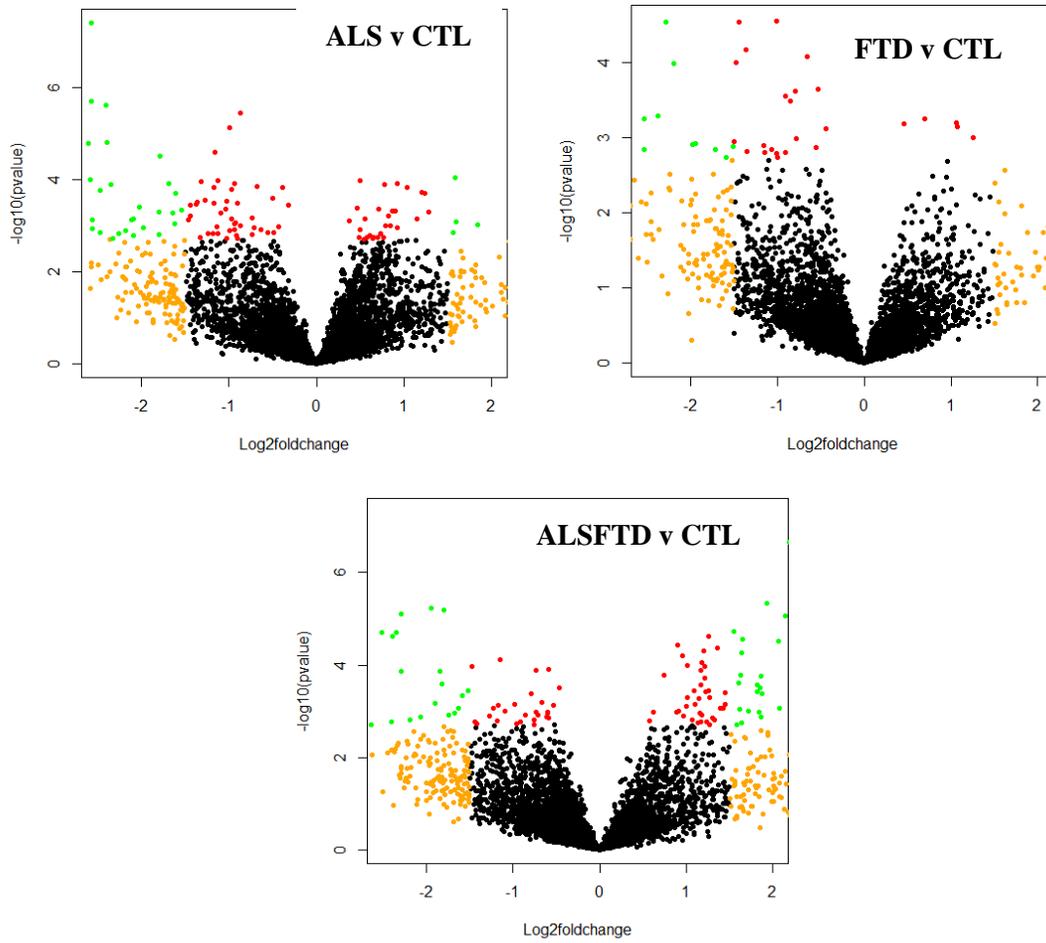
### **Global changes in the insoluble brain proteome**

To determine which of the identified proteins were preferentially enriched in ALS/FTD cases, protein comparison based on the extracted ion current (XIC) of identified peptides was conducted. Pairwise comparisons between controls and each disease group were conducted; these are represented in the volcano plots in **Figure 4.1**. Several proteins were significantly altered ( $\pm 1.5$ -fold,  $p < 0.05$ ) among the three pairwise comparisons between i.) controls vs. ALS, ii.) controls vs. ALS/FTD, and iii.) controls vs. FTD (**Figure 4.1**). The number of proteins significantly altered in ALS/FTD (n=245) was greater than that in ALS (n=169), in pairwise

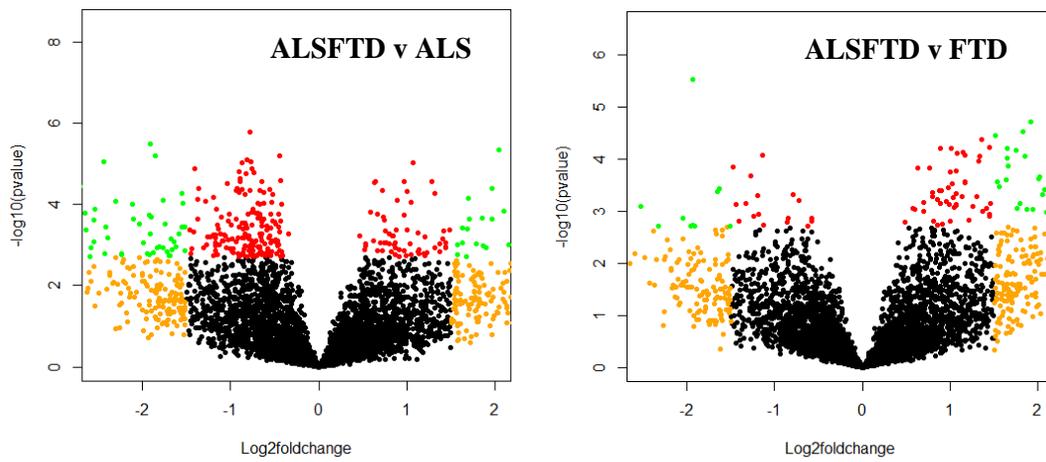
comparisons to controls. The number of significantly altered proteins in FTD compared to controls was lower (n=112). Additionally, pairwise comparisons across disease categories were conducted. These comparisons allow for a better understanding of the differences that are unique to the overlap syndrome. It is currently accepted that ALS/FTD represents an overlap of these disparate diseases, but not much is understood about its unique features. Not surprisingly, several proteins were significantly altered ( $\pm 1.5$ -fold,  $p < 0.05$ )

**Table 4.1. Spectral count and LFQ expression matrix for insoluble proteome  
(Supplemental file 2)**

A



B



**Figure 4.1. Significant changes in the detergent-insoluble brain proteome.** Volcano plots for ALS vs. CTL, ALSFTD vs CTL, FTD vs. CTL (A), ALS vs ALSFTD, and FTD vs. ALSFTD (B). Plots display the protein abundance ( $\log_2$  (fold change)) on the x-axis vs. the t-statistic ( $-\log_{10}$  (p-value)). Black dots represent proteins that remain unchanged, while the green dots represent proteins that are differentially enriched (right) or depleted (left) based on fold change difference  $\pm 1.5$  fold and p-value  $\leq 0.01$ .

in the pairwise comparison of ALS vs. ALSFTD and FTD vs. ALSFTD. The number of proteins significantly altered was greater in the comparison of ALSFTD to ALS (n=337) than the comparison of ALSFTD to FTD (n=29).

Accounting for multiple pairwise comparisons we used a more stringent significance cutoff of  $p \leq 0.01$  (since we ran 5 pairwise comparisons, Bonferroni correction for multiple pairwise comparisons (p threshold after correction:  $p=0.01$ )). The number of proteins significantly altered in ALSFTD (n=125) was larger compared to ALS (n=81), in pairwise comparisons to controls. The number of proteins significantly altered in FTD compared to controls was less (n=44). Several proteins were significantly altered ( $\pm 1.5$ -fold,  $p < 0.01$ ) in the pairwise comparison of ALS vs. ALSFTD and FTD vs. ALSFTD (**Figure 4.1**). The number of significant proteins was greater in the comparison of ALSFTD to ALS (n=196) than the comparison of ALSFTD to FTD (n=12).

In the pairwise comparison of ALS vs. controls annotation of proteins that were significantly altered showed an enrichment of RNA binding and RNA splicing/mRNA processing, while annotation of pairwise comparisons between ALSFTD vs controls showed enrichment for extracellular matrix structural constituents and negative regulation of protein metabolic processes. GO term enrichment analysis for proteins significantly altered in pairwise comparisons between FTD and controls identified structural constituents of ribosome, RNA binding proteins, and protein localization. Interestingly, analysis of proteins significantly altered in the pairwise comparisons between ALSFTD and FTD showed enrichment for extracellular matrix structural component proteins, which were significantly increased in ALSFTD cases

compared to FTD cases. Overall, these results show that the detergent-insoluble proteome displays robust differences across the ALS-FTD spectrum, and provides novel insights into disease associated alterations.

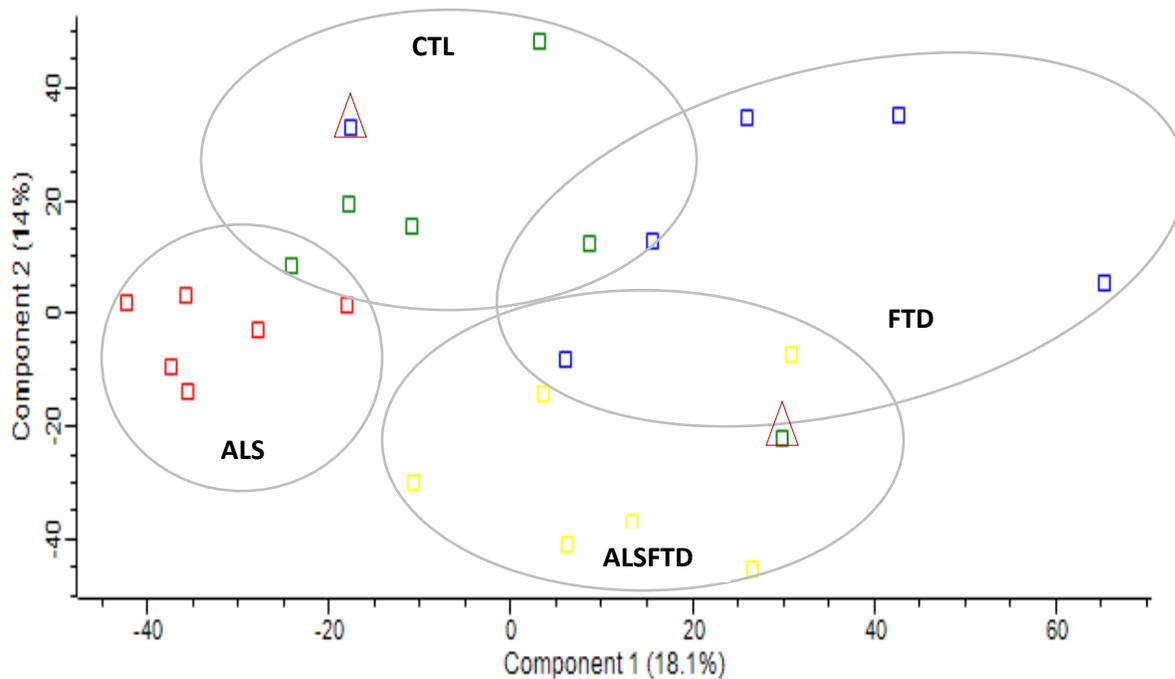
### **Segregation of clinical groups by proteomic signature**

All cases included in this study were selected based on clinical diagnoses and pathological diagnosis for FTD cases, based on TDP-43 being the major pathological overlap between ALS and FTD. Typical proteomic studies select cases based on similarities in neuropathology as correlations between neuropathology and proteomics are easier to imagine than clinical phenotypes and proteomic signatures. Interestingly, we found that the cases segregate out into groups based on clinical phenotypes using proteomic expression data, as shown in the multidimensional scaling plot showing the two principal components of the quantitative mass spectrometry data (**Figure 4.2**). Principal component analysis revealed an association with proteomic expression, as samples segregated into clusters which represented clinical groups. Notably, there were 2 outlier cases (1 control case and 1 FTD case). We see an overlap between FTD and ALSFTD clusters, but no overlap with ALSFTD and ALS. This is the first time to our knowledge that human postmortem samples have been used in this way and demonstrate proteomic differences that segregate the phenotypes.

### **Validation of detergent-insoluble LC-MC/MS data**

To independently validate the accuracy of LC-MS/MS generated data, we analyzed TDP-43 by immunohistochemical staining of pathological TDP (pTDP). Insoluble proteins typically

aggregate in disease and can be identified on pathological analysis. Immunohistochemistry serves as a

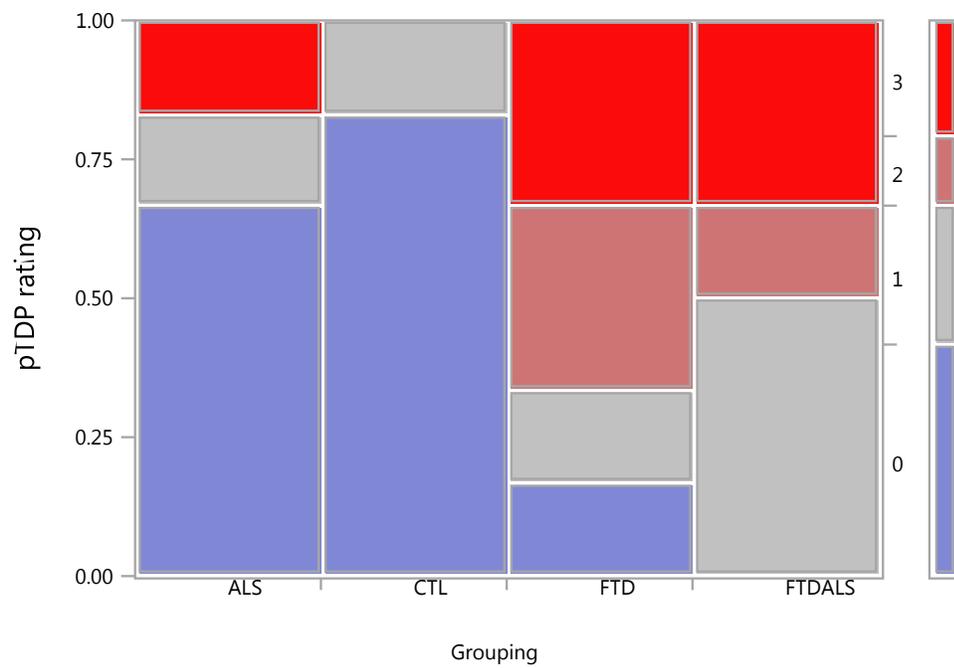


**Figure 4.2. Segregation of clinical groups by principal component analysis of detergent-insoluble LFQ proteomic data.** Each component on the PCA explains a larger amount of variance of the dataset. The x-axis represents the first principal component of the data, while the y-axis represents the second principal component. Clinical groups are indicated by colored squares (green: CTL, red: ALS, yellow: ALSFTD, blue: FTD). Segregation of cases based on clinical grouping is clearly demonstrated by the clusters, indicated by gray ovals. Outliers are shown in red triangles.

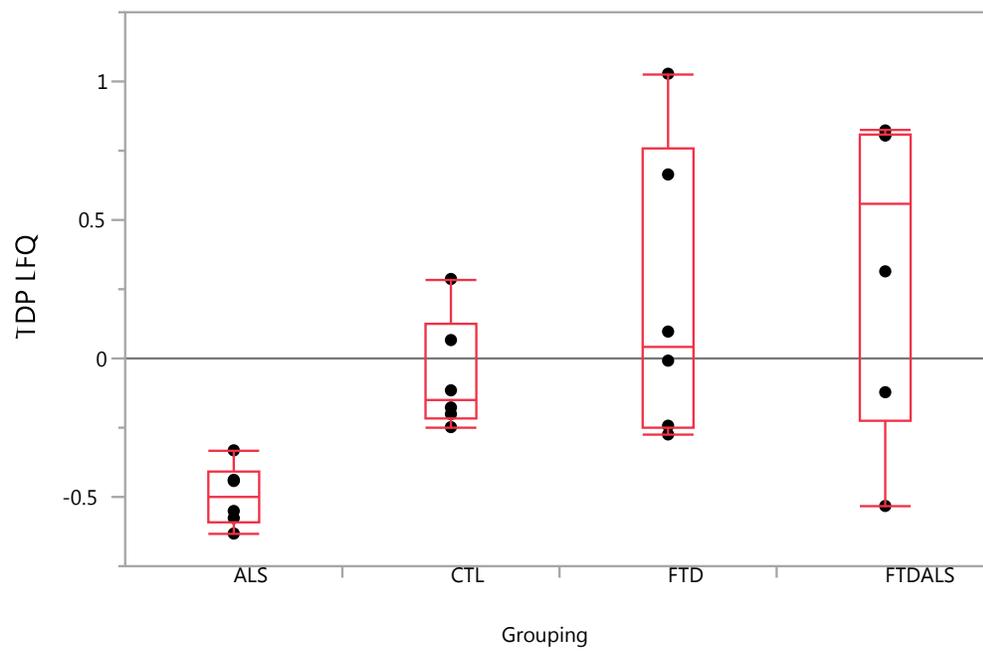
complimentary method to visualize alterations in proteins and their relationship to diagnosis along the TDP-43 disease spectrum. Frontal cortex tissue samples from FTD patients showed widespread TDP-43 neuropathology with increased immunoreactivity for phosphorylated TDP-43 (**Figure 4.3A**). Due to the limited number of cases in each group a fisher's exact test showed that the association between pTDP pathological score and clinical group was not significant( $p=0.0626$ ).

Label free quantification of TDP-43 across the cases showed a statistically significant increase in FTD and ALS/FTD cases compared to ALS and control cases (**Figure 4.3B**). This shows the increased sensitivity available with MS-based proteomic data compared to traditional immunohistochemical techniques and fits the expected pattern of pathology in the frontal cortex of FTD cases compared to ALS cases. Further we assessed the ratio of a C-terminal peptide extracted ion current (XIC) (representing abundance) to N-terminal peptide XIC, we found increased C-terminal TDP-43 peptide enrichment in FTD cases compared to ALS and controls (**Figure 4.3C**). Previous work has suggested that the C-terminal peptide is associated with potential toxicity of pathological TDP-43 (Neumann, Sampathu et al. 2006, Igaz, Kwong et al. 2008, Zhang, Xu et al. 2009, Tsuji, Arai et al. 2012). There was a high correlation between pTDP pathological rating and the C to N terminal TDP-43 ratio (**Figure 4.3D**;  $R^2=0.215$ ,  $p=0.0224$ ). This further supports the notion of using the C to N terminal TDP-43 ratio as a surrogate for neuropathology.

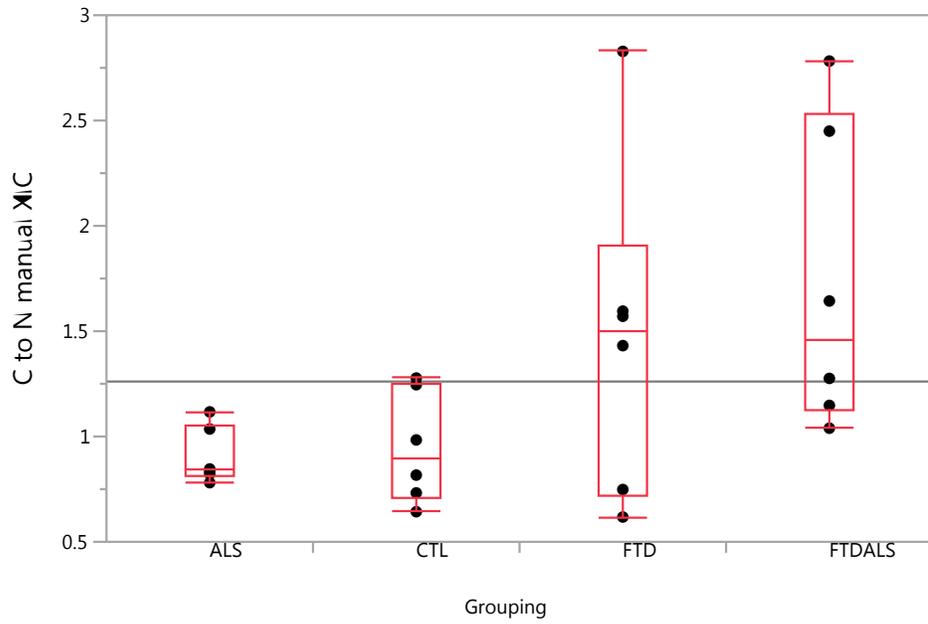
A.



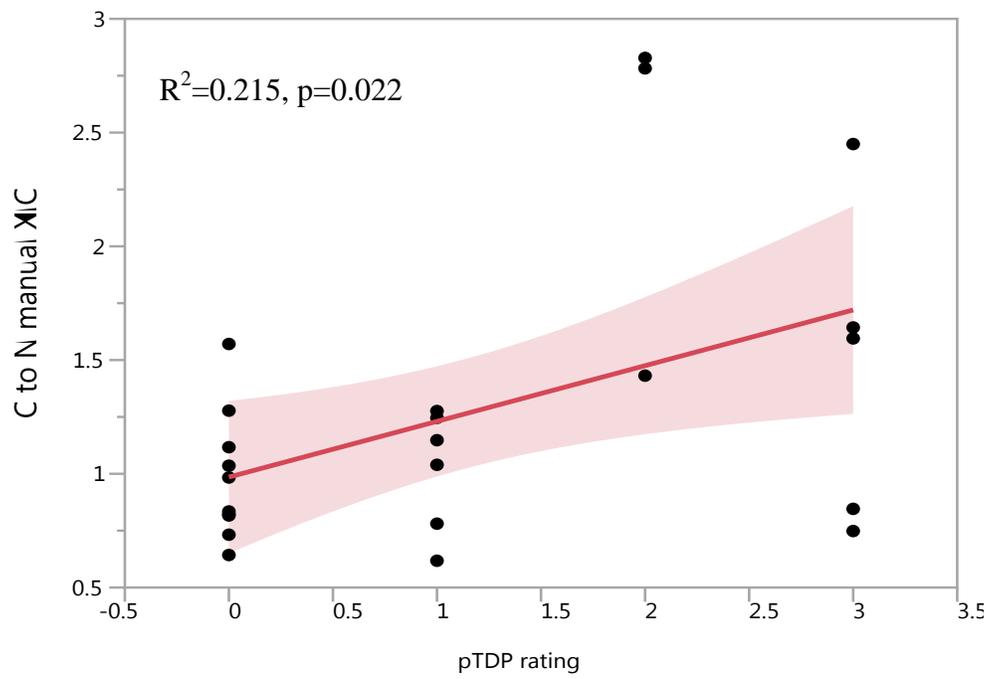
B.



C.



D.



**Figure 4.3. TDP-43 levels across samples used for detergent-insoluble proteomic analysis.**

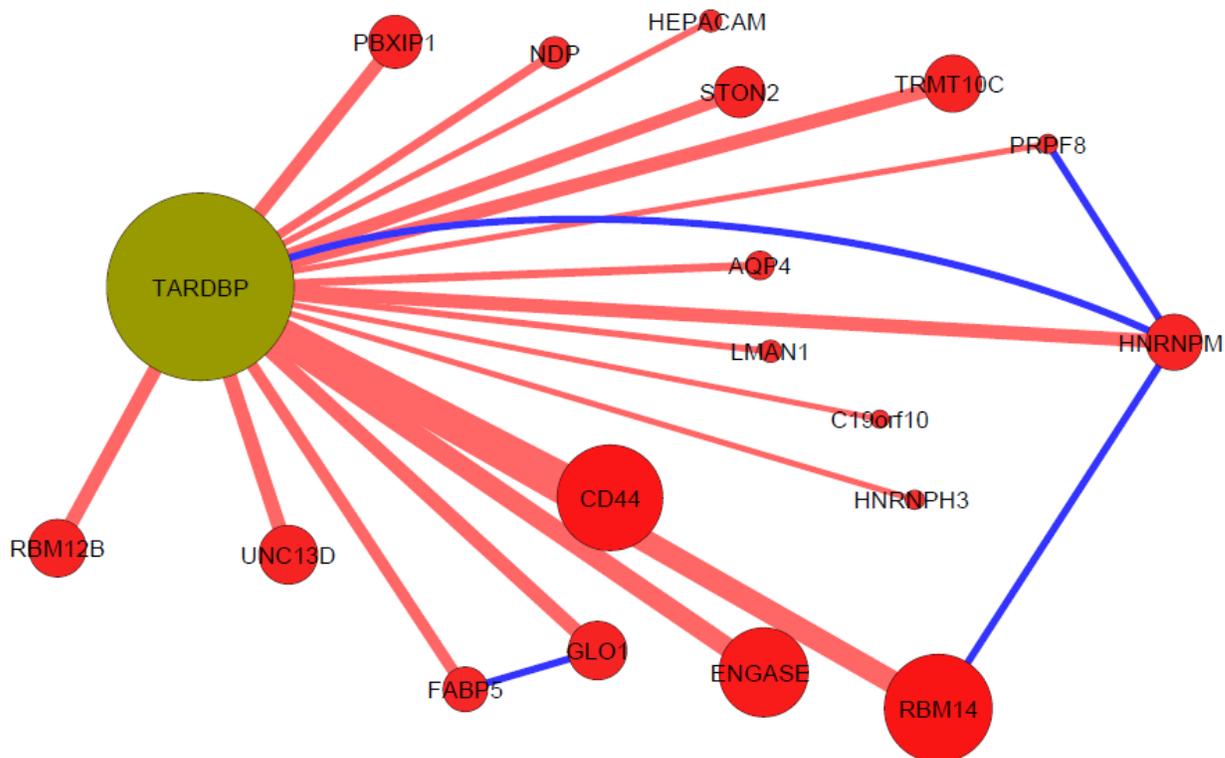
A. Mosaic plot displays disease groups along x-axis and phosphorylated TDP(pTDP) pathological score of 0-3 in different colors (0-blue, 1-gray, 2-muted red, and 3-red), with the percentage of cases with the respective score represented by the y-axis. B. Label-free quantification of TDP-43 across the cases is illustrated in this box plot graph which shows the mean TDP-43 across the groups. Individual case TDP-43 LFQ values are indicated by black dots. C. Box plots shown for C to N terminal TDP-43 peptide ratio, a case trait for these samples that was generated from manual XIC from LC-MS/MS analysis done on total homogenate. This shows enrichment of C-terminal TDP-43 in cases with FTD compared to ALS and CTL. D. Correlation plot showing significant correlation of pTDP rating to C to N-terminal TDP-43 ratio ( $R^2=0.215$ ,  $p=0.022$ ).

## **Defining proteins that correlate with TDP-43 levels in the detergent-insoluble brain proteome**

To assess proteomic changes across the ALS-FTD spectrum that correlate with TDP-43 in the insoluble proteome we performed a netscreen. A netscreen is carried out by determining the bi-weight mid-correlation (bicor) coefficient for all pairwise protein comparisons with TDP-43 LFQ, across all 24 individual cases. The top 20 proteins positively correlated with TARDBP LFQ were mostly RNA binding proteins; these are listed in **Table 4.2**. RBM14 was the most significantly correlated in the netscreen (bicor=0.74,  $p=3.38 \times 10^{-5}$ ), this RNA binding protein has been shown to act as a nuclear receptor coactivator that enhanced transcription (isoform 1) and a transcriptional repressor by modulating transcriptional activities of coactivators (isoform 2) (<http://www.genecards.org>). The next most correlated protein in the detergent-insoluble proteome to TDP-43 LFQ was CD44, typically considered an astrocytic protein. This cell surface glycoprotein has been shown to be involved in cell-cell interactions, cell adhesion, and migration; also, it has been linked to lymphocyte activation. UNC13D, a protein predominantly expressed in microglia, is involved in cytotoxic granule exocytosis in lymphocytes basically priming the immunological synapse (<http://www.genecards.org>), and was also significantly correlated to TDP-43 LFQ. Several hnRNPs (HNRNPM & HNRNPH3) were also significantly correlated to TDP-43 LFQ; these proteins have previously been associated with TDP-43 and ALS pathogenesis (Dammer, Fallini et al. 2012). Heterogeneous nuclear ribonucleoprotein M, HNRNPM, a pre-mRNA binding protein, has previously been shown to bind avidly to poly(G) and poly(U) RNA in vitro, and is involved in splicing (Kafasla, Patrino-Georgoula et al. 2002, Hovhannisyanyan and Carstens 2007, Marko, Leichter et al. 2014) .

<b>Protein ID (Gene  RefSeq ID)</b>	<b>Rank</b>	<b>Bicor</b>	<b>p-value</b>	<b>q-value (FDR)</b>
RBM14 NP_006319	1	0.74	3.38E-05	3%
CD44 NP_001189485	2	0.74	3.67E-05	3%
ENGASE NP_001036038	3	0.72	6.77E-05	3%
UNC13D NP_954712	4	0.69	2.11E-04	4%
GLO1 NP_006699	5	0.69	2.12E-04	4%
RBM12B NP_976324	6	0.69	2.20E-04	4%
TRMT10C NP_060289	7	0.68	2.24E-04	4%
HNRNPM NP_005959	8	0.68	2.34E-04	4%
PBXIP1 NP_065385	9	0.68	2.62E-04	4%
STON2 NP_149095	10	0.68	2.83E-04	4%
FABP5 NP_001435	11	0.67	3.56E-04	5%
NDP NP_000257	12	0.65	5.74E-04	5%
AQP4 NP_001641	13	0.65	6.49E-04	6%
LMAN1 NP_005561	14	0.64	8.24E-04	6%
HEPACAM NP_689935	15	0.64	8.29E-04	6%
PRPF8 NP_006436	16	0.63	9.17E-04	6%
HNRNPH3 NP_067676	17	0.63	9.19E-04	6%
C19orf10 NP_061980	18	0.63	9.63E-04	6%
CIRBP NP_001271	19	0.62	1.24E-03	6%
ZC3HAV1 NP_064504	20	0.62	1.26E-03	6%

**Table 4.2. Proteins most correlated with TDP-43(TARDBP) in the detergent -insoluble human brain proteome**



**Figure 4.4. Network analysis of protein co-aggregation based on correlation with TDP-43 in detergent insoluble proteome.** TDP-43 (TARDBP) (large yellow node) correlates with several proteins measured in the insoluble proteome from the 24 cases analyzed and quantified by label-free MS quantification. In the diagram, the node size is proportional to significance (Z score from netscreen), thickness of red connecting line is proportional to the bicor correlation value for each pairwise interaction, and blue lines indicate protein-protein interactions. Interactions are derived from BioGrid database (version 3.4.147). The top positively correlated proteins, with p-value  $\leq 0.001$  are shown in this diagram (n=18).

Further, RBM12B an RNA binding motif protein, and, HEPACAM, hepatic and glial cell adhesion molecule, an astrocytic protein involved in regulating cell motility and cell matrix interactions were among the top significantly correlated proteins. PRPF8 (PRP8 pre-mRNA processing factor 8 homolog) was also a top correlated protein to TDP-43 LFQ; this protein has been shown to mediate the ordered assembly of spliceosomal proteins and snRNAs such that splicing can occur. Finally, cold inducible RNA binding protein (CIRBP) is an mRNA binding protein that plays a protective role in the stress response by stabilizing transcripts of genes involved in cell survival; it had been shown to act as a translational activator by binding specifically to the 3'-untranslated regions (3'-UTRs) of stress-responsive transcripts (<http://www.genecards.org>). Figure 4.4 is a Network analysis of protein co-aggregation based on correlation with TDP-43 in detergent insoluble proteome.t represents the known connections/interactions between the most positively correlated proteins to TDP-43 LFQ.

## **Discussion**

Previously, LC-MS/MS has been utilized for the quantitative analysis of the detergent-insoluble brain proteome in FTLN cases (Seyfried, Gozal et al. 2012). From this analysis, novel RNA-binding proteins were identified as enriched and mislocalized in FTLN cases compared to controls. Additionally, this unbiased approach has been applied to uncover proteins involved in AD pathogenesis and more clearly delineate molecular aspects that were previously unknown (Hales, Dammer et al. 2016). Similarly, other groups have used LC-MS/MS for biomarker discovery and to uncover disease pathogenesis in ALS (Chen, Liu et al. 2016)(Chen et al. 2016)(Basso, Samengo et al. 2009, Collins, An et al. 2015, Stalekar, Yin et al. 2015, Blokhuis, Koppers et al. 2016, Engelen-Lee, Blokhuis et al. 2016, Li, Collins et al. 2016). However, no

group has examined the insoluble proteome across the TDP-43 spectrum of diseases. This is important because it will give us a better understanding of the basis of the overlap of these seemingly disparate diseases. Therefore, we compared the insoluble proteome across TDP-43 proteinopathies from patients with ALS, ALSFTD, and FTD to determine differential expression and protein co-expression of insoluble proteins to get a better understanding of this spectrum of disease.

In this analysis, we leveraged label-free mass spectrometry based proteomics to quantify 4,647 detergent-insoluble proteins across the frontal cortex of 24 individual cases representing control, ALS, ALSFTD, and FTD. As anticipated, TDP-43 was significantly increased in cases with FTD, and this significantly correlated with pathological scores of pTDP. Differential expression analysis was performed to identify differential protein expression across the ALS-FTD disease spectrum. Interestingly, many of the significantly altered proteins in comparison of ALS to controls were RNA binding proteins. This is what others have also shown in the literature, though from motor regions. Cellular components enriched for paraspeckles and mitochondria were enriched in ALS brains compared to controls. RNA binding proteins, proteins involved in protein localization to the ER membrane, and structural constituents of the ribosome were overrepresented in the comparison of FTD cases to controls; cellular enrichment analysis showed most of these proteins were ribosomal, associated through the ribonucleoprotein complex. We had expected to see synaptic or neuronal protein differences based on underlying pathology in the frontal lobe of FTD patients. Further, proteins significantly altered in ALSFTD compared to controls were mostly extracellular matrix organization proteins. This was surprising, as we expected to see an intermediate picture or overlap of differentially expressed proteins in

ALSFTD vs CTL compared to ALS vs CTL and FTD vs. CTL. This provides evidence that patients with the overlapping syndrome may possess a unique disease entity that arises from a different mechanistic route but converges on similar clinical and pathological features. This shakes the assumption that the overlap simply represents the intersection of ALS and FTD pathway alterations.

In pairwise comparisons of ALS to ALSFTD and FTD to ALSFTD there were fewer significantly altered proteins when comparing patients with FTD than those with ALS. As the frontal cortex was used for this analysis this was somewhat expected. Proteins significantly altered in the ALSFTD to ALS comparison represented proteins with kinase activity and those involved in synaptic signaling; enrichment analysis uncovered dendrite and postsynaptic specialization as significant GO cellular components. Meanwhile, the few proteins significantly altered in the comparison of FTD to FTDALS were mostly structural proteins, components of the extracellular matrix.

To investigate protein co-aggregation activities linked to the hallmark ALS-FTD pathology of TDP-43, pairwise correlation coefficients were calculated for each quantified protein to TDP-43 LFQ, across all cases. The most correlated proteins to TDP-43 LFQ were RNA binding proteins, notably RBM14. RBM14 is a paraspeckle RNA binding protein that has previously been shown to co-aggregate with TDP-43 in cell lines (Dammer, Fallini et al. 2012). Other proteins with high correlation to TDP were proteins known to be involved in inhibiting transcriptional activation potential through preventing the PBX1 homeodomain protein from binding to DNA (PBXIP1), and cold-inducible mRNA binding protein (CIRB) which plays a role in the stress response by

stabilizing transcripts of genes involved in cell survival by binding to the 3'-untranslated regions (3'-UTRs) of stress responsive transcripts to act as a translational activator. CIRB promotes assembly of stress granules (SGs), when overexpressed. hnRNPM, another protein highly correlated to TDP-43 LFQ across the cases, is a splicing factor that has been shown to interact with TAF15 (a member of the FET protein family- a group of DNA and RNA binding proteins). Abnormal RNA processing has been previously implicated in ALS pathobiology as factors that contribute to protein dysfunction (Donnelly, Grima et al. 2014). Another category of proteins that were highly correlated to TDP-43 LFQ are proteins enriched in microglial and astrocytic cell types; these include CD44 (astrocytic), UNC13D (microglial), HEPACAM (astrocytic), ENGASE (microglial). This correlation may suggest increased inflammation correlated to increased TDP-43 along the ALS-FTD disease spectrum. Inflammation has previously been implicated in ALS-FTD through microglial involvement and correlation with disease progression (Brettschneider, Toledo et al. 2012, Radford, Morsch et al. 2015).

A major limitation of this study was the proteome coverage; however, this is a consistent limitation of all shot-gun proteomic studies (Gstaiger and Aebersold 2009). Other limitations include the limited sample size. At the start of this project the equipment we used limited the number of cases we could run. Another limitation is the age range of the cases included in the analysis. Ideally, we would have wanted our control cases to be around the same age as the disease cases, but we were limited to using the material available in the Brain bank. Nevertheless, mass spectrometry based proteomics allows for the identification and quantitation of different proteomic profiles across the ALS-FTD continuum. Thus, the systemic analysis of

protein alterations presents a new level of data to use in understanding how specific phenotypes arise.

Our findings highlight the utility of label-free mass spectrometry based proteomics to identify disease-specific proteins that are common across TDP-43 proteinopathies and those that are unique to cases with dementia. Using differential expression analysis, we get a better sense of proteins altered in each disease group, and can start to infer pathways that may be involved in disease pathogenesis. Further characterization of differentially expressed proteins identified from this analysis is needed to tease apart their role in disease pathogenesis. However, this study serves as a crucial first step in associating protein alterations in brains from the detergent-insoluble proteome of patients along the ALS-FTD spectrum to clinical phenotypes. This study serves to inform the field by assessing correlation of protein changes in disease to pathological measures, and with this identified previously known and novel pathways implicated in this disease spectrum. Mechanistic studies of these novel targets could inform our understanding of pathways involved in these co-occurring diseases. Further analysis of brain proteomes will increase our understanding of the pathways linking these seemingly disparate diseases.

**Chapter 5: A proteomic network approach across the ALS-FTD disease spectrum resolves clinical phenotypes and genetic vulnerability in human brain**

## **Introduction**

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are clinically distinct neurodegenerative diseases that are connected by genetic and pathological overlap (Fecto and Siddique 2011, Ferrari, Kapogiannis et al. 2011, Achi and Rudnicki 2012). ALS patients present with muscle weakness and spasticity associated with degeneration of motor neurons in the motor cortex, brainstem, and spinal cord that ultimately leads to death. In contrast, patients with FTD display cognitive dysfunction associated with degeneration of neurons in the frontal and temporal lobes of the brain. Despite being clinically distinct, 15% of individuals presenting with FTD also have ALS, whereas 30% of individuals with ALS will develop FTD (Lomen-Hoerth 2011). This implies that these two neurodegenerative diseases are part of a shared clinical spectrum.

In addition to their clinical overlap, most cases of ALS and FTD display pathological accumulation of TAR-DNA binding protein (TDP-43), a ubiquitously expressed nuclear DNA/RNA binding protein that is cleaved, phosphorylated, and aggregated in the cytoplasm in disease (Neumann, Sampathu et al. 2006). Ninety-seven percent of ALS cases display phosphorylated TDP-43 pathology in the brain and/or spinal cord, while 50% of FTD cases display this pathology throughout the brain (Radford, Morsch et al. 2015); defining these diseases as TDP proteinopathies. Many individuals with ALS and FTD also share a positive family history of disease (Fong, Karydas et al. 2012). The largest proportion of inherited cases (40% ALS and 25% FTD) are caused by hexanucleotide G<sub>4</sub>C<sub>2</sub> repeat expansions in the C9orf72 gene, which notably was identified from families with co-occurring ALS and FTD (ALS/FTD) (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011, Majounie, Renton et

al. 2012, Radford, Morsch et al. 2015). Although the function of the C9orf72 protein is not yet known, there are several theories regarding how the C9orf72 mutation leads to ALS and FTD (Zhang, Iyer et al. 2012, Farg, Sundaramoorthy et al. 2014). For example, loss of C9orf72 protein expression is thought to inhibit autophagy and promote neuroinflammation (Farg, Sundaramoorthy et al. 2014, Webster, Smith et al. 2016), whereas expression of C9orf72 gene products could cause toxicity via nuclear sense and antisense repeat-containing RNAs that sequester RNA binding proteins (abnormal RNA metabolism) or by noncanonical repeat-associated non-ATG (RAN) translation of dipeptide repeat proteins that aggregate and block nuclear pores (Gitler and Tsuiji 2016). Notably, this genetic mutation bridges these two diseases that clinical and pathological overlaps have previously connected. However, the underlying molecular basis of the ALS-FTD clinical and pathological spectrum is not well established. It is also unclear why patients with the same C9orf72 genetic expansion get either or both of these disparate diseases. Using our knowledge of the shared clinical, pathological and genetic features characterizing ALS and FTD, a systems level proteomic analysis of both sporadic and genetic (C9orf72) cases comprising the ALS-FTD spectrum was conducted to determine common and distinct pathways that contribute to the onset and development of dementia.

Co-expression network analysis has been used to define modules of co-expressed genes or proteins linked to specific cell types, organelles, and biological pathways (Miller, Oldham et al. 2008, Oldham 2014, Seyfried, Dammer et al. 2017). Assessing co-expression of proteins within samples and relating co-expression modules to clinical and pathological traits can be defined utilizing weighted co-expression network analysis (WGCNA), where the most centrally connected proteins in a module act as key drivers (Zhang and Horvath 2005, Oldham, Konopka

et al. 2008). We recently reported the first large-scale proteomic and transcriptomic multi-network analysis in human post-mortem brain from both asymptomatic and symptomatic Alzheimer's disease (AD) (Seyfried, Dammer et al. 2017). This work revealed that several protein-driven processes related to cognitive decline are distinct from networks in human AD transcriptome. Moreover, analysis of the proteome is particularly relevant since neurodegenerative diseases are collectively viewed as proteinopathies defined by their association with the aggregation and accumulation of misfolded proteins (Dugger and Dickson 2017).

Here, we report the first unbiased proteomic analysis of 51 post-mortem cortical tissues from clinically characterized ALS, FTD, ALS/FTD, and healthy disease controls. A subset of C9orf72 expansion positive (C9Pos) cases was also included. WGCNA revealed 15 modules of co-expressed proteins, 8 of which were significantly different across the ALS-FTD disease spectrum. These included modules associated with RNA binding proteins, synaptic transmission, inflammation and cell-type specificity (neuronal, microglial and astrocytic), that showed strong correlation with TDP-43 pathology and cognitive dysfunction. Compared to sporadic ALS patients, C9Pos ALS cases showed increased levels of proteins associated with astrocytic and microglial markers, which supports a hypothesis that links the C9orf72 mutation to neuroinflammation.

## **Results**

### *Proteomic signature of human brain classifies cases by clinical phenotypes*

This study offers an in-depth analysis of protein changes in the frontal cortex of patients across the ALS-FTD disease spectrum which resulted in the final quantification of 2,612 protein groups

mapping to 2,536 unique gene symbols across all samples (**Table 5.1**). Our goal was to compare ALS patients with and without clinical dementia, so our experimental case selection approach grouped samples based on clinical phenotype. As FTD is clinically heterogeneous, we limited that group to include only individuals with the pathological diagnosis of frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP), which accounts for the majority of overlap with ALS and FTD (Neumann, Rademakers et al. 2009). The frontal cortex was chosen because it was a likely site for discriminating proteomic differences in ALS patients with and without dementia, and represents an area in which TDP-43 pathological burden has been mapped in FTD patients (Brettschneider, Del Tredici et al. 2014). ANOVA comparisons of each of the disease groups and controls yielded subsets of significantly altered proteins (**Table 5.1**). Interestingly, the number of significantly altered proteins increased when comparing controls to ALS, ALSFTD, and FTD, which is likely due to the presence of clinical dementia and the associated pathological burden in the frontal cortex (**Figure 5.1A**). For supervised clustering analysis, we selected 165 proteins that had at least 2 ANOVA Tukey pairwise comparisons of high significance ( $p < 0.01$ ) among the 6 possible comparisons among the clinical groups. The over-represented gene ontology (GO) terms within these 165 information rich proteins included cytoplasmic vesicle part, clathrin coat, and plasma membrane (**Table 5.2**). Using these differentially expressed protein signatures of disease, we used multidimensional scaling (MDS) to stratify the relatedness of individual cases within a dataset, which revealed that clinical phenotypes indeed associate with proteomic signatures, as samples segregated into clusters representing clinical groups (**Figure 5.1B**). Also, ALS/FTD cases distributed between the ALS and the FTD clusters, supporting an underlying molecular, biological spectrum designated by protein expression that defines the clinical spectrum. Thus, this analysis demonstrates that

protein signatures of FTD-ALS clinical phenotypes are related to differences in molecular pathways in the frontal cortex.

#### *Protein co-expression network analysis*

Protein co-expression networks reflect relationships between protein pathways, cell types, and physically interacting proteins within modules. Network analyses revealed 15 modules of strongly co-expressed groups of proteins (**Table 5.1**). Each module is defined by an eigenprotein, the most representative weighted protein expression pattern across samples for a group of co-expressed proteins (Seyfried, Dammer et al. 2017). Eight of these 15 modules, identified by numbers that correspond to a color ordered 1 to 15, with M1 (largest, 548 proteins) to M15 (smallest, 21 proteins), showed different patterns of co-expression across the 4 clinical groups. As expected, functional annotation of frontal cortex modules classified the proteome into modules associated with specific gene ontologies and brain cell types among other biological sources of co-expression (Gaiteri, Ding et al. 2014). Of the significant modules, 4 (M2, M6, M9, M10) showed increased expression in FTD cases compared to ALS and control cases (**Figure 5.2A**). These modules were enriched for protein transport, inflammation, and RNA binding proteins (**Table 5.2**). One of the significant modules, M15, enriched for blood microparticles and circulating immunoglobulin complexes, showed increased protein expression in all disease groups compared to controls, consistent with a common mechanism of blood brain barrier breakdown in neurodegenerative diseases (**Figure 5.2A**) (Carvey, Hendey et al. 2009, Seyfried, 2017 #12)). The remaining 3 significant modules (M1, M3, M8) showed a decrease in protein expression in FTD cases compared to control and ALS cases; M1 and M8 were enriched in synaptic and neuronal proteins, while M3 was enriched with mitochondrial proteins (**Figure 5.2A, Table 5.2**). Moreover, several modules showed significant correlation with clinical and

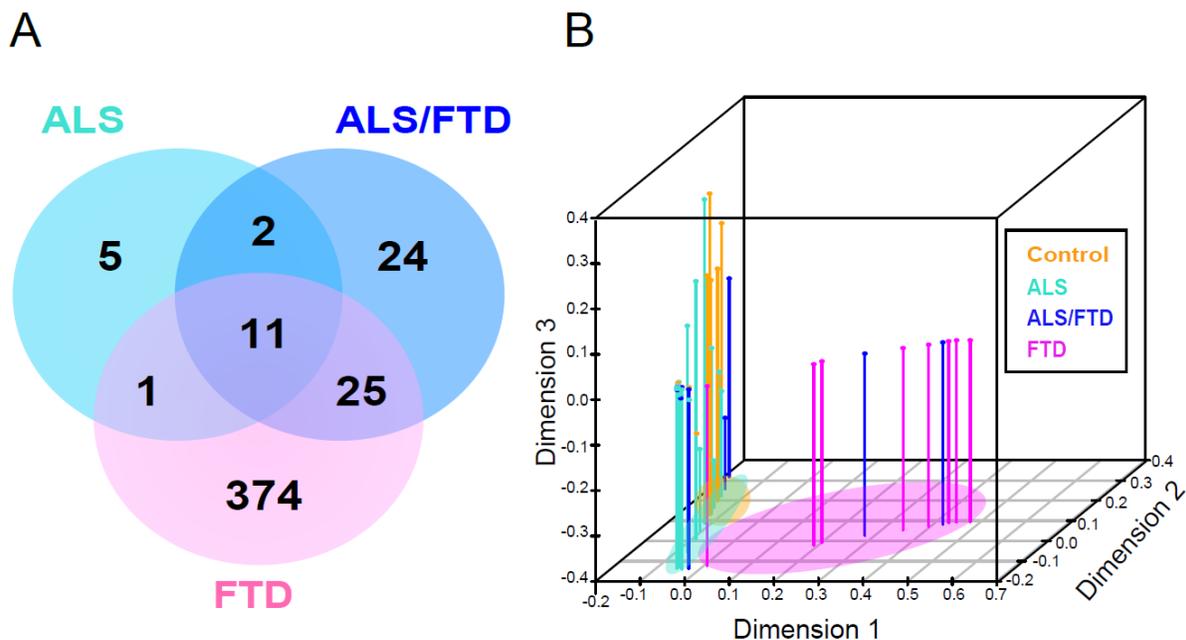
pathological sample traits. Pathological sample traits characterize TDP-43 pathology, which is represented by the abundance of phosphorylated TDP-43 cytoplasmic inclusions in the frontal cortex (defined by the pTDP score) or label-free protein quantification (LFQ) of TDP-43 (**Figure 5.3**). For example, M5, a module enriched with extracellular matrix and astrocyte proteins correlated closely to clinical and pathological traits (**Figure 5.4A**) and M14, a module enriched with microtubule proteins with dynactin as a hub protein, and optineurin, genes previously implicated in ALS and FTD (Fecto and Siddique 2011), correlated with pathological TDP-43 scores and TDP-43 levels. Additionally, M7, enriched with proteins involved in protein transport, correlated with TDP-43 pathology (**Figure 5.4A**).

As expected for the frontal cortex of ALS cases, where TDP-43 pathology is typically not found, and no cognitive decline phenotype exists, module eigenprotein expression for ALS patients without dementia was similar to controls. There was little overlap between controls and either FTD or ALS/FTD. Notably, we found differences in eigenprotein expression between ALS/FTD and FTD alone. Namely, the eigenprotein expression for many of the modules representing the ALS/FTD group was intermediate between that of the strictly ALS and strictly FTD clinical groups, supporting the existence of a molecular spectrum of disease. Mapping of differentially expressed proteins to co-expression modules in cases with ALS/FTD and FTD further supports an underlying molecular continuum (**Figure 5.2B**). There was an enrichment of differentially expressed proteins altered in pairwise comparisons comparing ALS/FTD or FTD patients to controls that were negatively correlated to neuronal modules (M1 and M8) and positively correlated to M2, M6, M9, and M10 modules. Additionally, the proteomic network generated in this study was compared to a previously generated proteomic network created from frontal

cortex samples of Emory ALS, AD, and PD cases (Seyfried, Dammer et al. 2017). Protein co-expression networks were highly preserved when comparing both protein networks (**Figure 5.5A**) and over representation analysis showed that all 15 modules generated in the current study had at least one cognate module within the previously generated Emory brain protein network (**Figure 5.5B**). The consistency between these independent datasets provides confidence in the networks generated in this study, and provides opportunity to expose meaningful relationships along the ALS-FTD disease spectrum by relating clinical and pathological sample traits to modules.

**Table 5.1. Differential and co-expression protein analysis (Supplemental file 3)**

**Table 5.2. Gene ontology (GO) term enrichment for co-expression network (Supplemental file 4)**



**Figure 5.1. Segregation of clinical ALS-FTD phenotypes by proteomic signatures**

**A.** Venn diagram showing a total of 442 unique proteins determined to be significantly altered by ANOVA followed by Tukey's post-hoc test ( $p \leq 0.01$ ) among 3 pairwise comparisons (i) ALS, (ii.) ALS/FTD, and (iii) FTD versus control cases. **B.** Supervised hierarchical clustering of 165 significant proteins altered in the frontal cortex across clinical phenotypes (CTL, ALS, ALS-FTD, and FTD) by criteria described in methods is shown on by a 3-dimensional multi-dimensional scaling (MDS) analysis. Each dimension on the plot explains a larger proportion of variance of the dataset. Clinical groups are indicated by colors (orange-controls, light blue-ALS, dark blue-FTD/ALS, and pink-FTD). Segregation of cases based on clinical grouping is indicated by colored ovals.

ALSFTD vs CTL compared to ALS vs CTL and FTD vs. CTL. This provides evidence that patients with the overlapping syndrome may possess a unique disease entity that arises from a different mechanistic route but converges on similar clinical and pathological features. This shakes the assumption that the overlap simply represents the intersection of ALS and FTD pathway alterations.

In pairwise comparisons of ALS to ALSFTD and FTD to ALSFTD there were fewer significantly altered proteins when comparing patients with FTD than those with ALS. As the frontal cortex was used for this analysis this was somewhat expected. Proteins significantly altered in the ALSFTD to ALS comparison represented proteins with kinase activity and those involved in synaptic signaling; enrichment analysis uncovered dendrite and postsynaptic specialization as significant GO cellular components. Meanwhile, the few proteins significantly altered in the comparison of FTD to FTDALS were mostly structural proteins, components of the extracellular matrix.

To investigate protein co-aggregation activities linked to the hallmark ALS-FTD pathology of TDP-43, pairwise correlation coefficients were calculated for each quantified protein to TDP-43 LFQ, across all cases. The most correlated proteins to TDP-43 LFQ were RNA binding proteins, notably RBM14. RBM14 is a paraspeckle RNA binding protein that has previously been shown to co-aggregate with TDP-43 in cell lines (Dammer, Fallini et al. 2012). Other proteins with high correlation to TDP were proteins known to be involved in inhibiting transcriptional activation potential through preventing the PBX1 homeodomain protein from binding to DNA (PBXIP1), and cold-inducible mRNA binding protein (CIRB) which plays a role in the stress response by

stabilizing transcripts of genes involved in cell survival by binding to the 3'-untranslated regions (3'-UTRs) of stress responsive transcripts to act as a translational activator. CIRB promotes assembly of stress granules (SGs), when overexpressed. hnRNPM, another protein highly correlated to TDP-43 LFQ across the cases, is a splicing factor that has been shown to interact with TAF15 (a member of the FET protein family- a group of DNA and RNA binding proteins). Abnormal RNA processing has been previously implicated in ALS pathobiology as factors that contribute to protein dysfunction (Donnelly, Grima et al. 2014). Another category of proteins that were highly correlated to TDP-43 LFQ are proteins enriched in microglial and astrocytic cell types; these include CD44 (astrocytic), UNC13D (microglial), HEPACAM (astrocytic), ENGASE (microglial). This correlation may suggest increased inflammation correlated to increased TDP-43 along the ALS-FTD disease spectrum. Inflammation has previously been implicated in ALS-FTD through microglial involvement and correlation with disease progression (Brettschneider, Toledo et al. 2012, Radford, Morsch et al. 2015).

A major limitation of this study was the proteome coverage; however, this is a consistent limitation of all shot-gun proteomic studies (Gstaiger and Aebersold 2009). Other limitations include the limited sample size. At the start of this project the equipment we used limited the number of cases we could run. Another limitation is the age range of the cases included in the analysis. Ideally, we would have wanted our control cases to be around the same age as the disease cases, but we were limited to using the material available in the Brain bank. Nevertheless, mass spectrometry based proteomics allows for the identification and quantitation of different proteomic profiles across the ALS-FTD continuum. Thus, the systemic analysis of

protein alterations presents a new level of data to use in understanding how specific phenotypes arise.

Our findings highlight the utility of label-free mass spectrometry based proteomics to identify disease-specific proteins that are common across TDP-43 proteinopathies and those that are unique to cases with dementia. Using differential expression analysis, we get a better sense of proteins altered in each disease group, and can start to infer pathways that may be involved in disease pathogenesis. Further characterization of differentially expressed proteins identified from this analysis is needed to tease apart their role in disease pathogenesis. However, this study serves as a crucial first step in associating protein alterations in brains from the detergent-insoluble proteome of patients along the ALS-FTD spectrum to clinical phenotypes. This study serves to inform the field by assessing correlation of protein changes in disease to pathological measures, and with this identified previously known and novel pathways implicated in this disease spectrum. Mechanistic studies of these novel targets could inform our understanding of pathways involved in these co-occurring diseases. Further analysis of brain proteomes will increase our understanding of the pathways linking these seemingly disparate diseases.

**Chapter 5: A proteomic network approach across the ALS-FTD disease spectrum resolves clinical phenotypes and genetic vulnerability in human brain**

## **Introduction**

Amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD) are clinically distinct neurodegenerative diseases that are connected by genetic and pathological overlap (Fecto and Siddique 2011, Ferrari, Kapogiannis et al. 2011, Achi and Rudnicki 2012). ALS patients present with muscle weakness and spasticity associated with degeneration of motor neurons in the motor cortex, brainstem, and spinal cord that ultimately leads to death. In contrast, patients with FTD display cognitive dysfunction associated with degeneration of neurons in the frontal and temporal lobes of the brain. Despite being clinically distinct, 15% of individuals presenting with FTD also have ALS, whereas 30% of individuals with ALS will develop FTD (Lomen-Hoerth 2011). This implies that these two neurodegenerative diseases are part of a shared clinical spectrum.

In addition to their clinical overlap, most cases of ALS and FTD display pathological accumulation of TAR-DNA binding protein (TDP-43), a ubiquitously expressed nuclear DNA/RNA binding protein that is cleaved, phosphorylated, and aggregated in the cytoplasm in disease (Neumann, Sampathu et al. 2006). Ninety-seven percent of ALS cases display phosphorylated TDP-43 pathology in the brain and/or spinal cord, while 50% of FTD cases display this pathology throughout the brain (Radford, Morsch et al. 2015); defining these diseases as TDP proteinopathies. Many individuals with ALS and FTD also share a positive family history of disease (Fong, Karydas et al. 2012). The largest proportion of inherited cases (40% ALS and 25% FTD) are caused by hexanucleotide G<sub>4</sub>C<sub>2</sub> repeat expansions in the C9orf72 gene, which notably was identified from families with co-occurring ALS and FTD (ALS/FTD) (DeJesus-Hernandez, Mackenzie et al. 2011, Renton, Majounie et al. 2011, Majounie, Renton et

al. 2012, Radford, Morsch et al. 2015). Although the function of the C9orf72 protein is not yet known, there are several theories regarding how the C9orf72 mutation leads to ALS and FTD (Zhang, Iyer et al. 2012, Farg, Sundaramoorthy et al. 2014). For example, loss of C9orf72 protein expression is thought to inhibit autophagy and promote neuroinflammation (Farg, Sundaramoorthy et al. 2014, Webster, Smith et al. 2016), whereas expression of C9orf72 gene products could cause toxicity via nuclear sense and antisense repeat-containing RNAs that sequester RNA binding proteins (abnormal RNA metabolism) or by noncanonical repeat-associated non-ATG (RAN) translation of dipeptide repeat proteins that aggregate and block nuclear pores (Gitler and Tsuiji 2016). Notably, this genetic mutation bridges these two diseases that clinical and pathological overlaps have previously connected. However, the underlying molecular basis of the ALS-FTD clinical and pathological spectrum is not well established. It is also unclear why patients with the same C9orf72 genetic expansion get either or both of these disparate diseases. Using our knowledge of the shared clinical, pathological and genetic features characterizing ALS and FTD, a systems level proteomic analysis of both sporadic and genetic (C9orf72) cases comprising the ALS-FTD spectrum was conducted to determine common and distinct pathways that contribute to the onset and development of dementia.

Co-expression network analysis has been used to define modules of co-expressed genes or proteins linked to specific cell types, organelles, and biological pathways (Miller, Oldham et al. 2008, Oldham 2014, Seyfried, Dammer et al. 2017). Assessing co-expression of proteins within samples and relating co-expression modules to clinical and pathological traits can be defined utilizing weighted co-expression network analysis (WGCNA), where the most centrally connected proteins in a module act as key drivers (Zhang and Horvath 2005, Oldham, Konopka

et al. 2008). We recently reported the first large-scale proteomic and transcriptomic multi-network analysis in human post-mortem brain from both asymptomatic and symptomatic Alzheimer's disease (AD) (Seyfried, Dammer et al. 2017). This work revealed that several protein-driven processes related to cognitive decline are distinct from networks in human AD transcriptome. Moreover, analysis of the proteome is particularly relevant since neurodegenerative diseases are collectively viewed as proteinopathies defined by their association with the aggregation and accumulation of misfolded proteins (Dugger and Dickson 2017).

Here, we report the first unbiased proteomic analysis of 51 post-mortem cortical tissues from clinically characterized ALS, FTD, ALS/FTD, and healthy disease controls. A subset of C9orf72 expansion positive (C9Pos) cases was also included. WGCNA revealed 15 modules of co-expressed proteins, 8 of which were significantly different across the ALS-FTD disease spectrum. These included modules associated with RNA binding proteins, synaptic transmission, inflammation and cell-type specificity (neuronal, microglial and astrocytic), that showed strong correlation with TDP-43 pathology and cognitive dysfunction. Compared to sporadic ALS patients, C9Pos ALS cases showed increased levels of proteins associated with astrocytic and microglial markers, which supports a hypothesis that links the C9orf72 mutation to neuroinflammation.

## **Results**

### *Proteomic signature of human brain classifies cases by clinical phenotypes*

This study offers an in-depth analysis of protein changes in the frontal cortex of patients across the ALS-FTD disease spectrum which resulted in the final quantification of 2,612 protein groups

mapping to 2,536 unique gene symbols across all samples (**Table 5.1**). Our goal was to compare ALS patients with and without clinical dementia, so our experimental case selection approach grouped samples based on clinical phenotype. As FTD is clinically heterogeneous, we limited that group to include only individuals with the pathological diagnosis of frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP), which accounts for the majority of overlap with ALS and FTD (Neumann, Rademakers et al. 2009). The frontal cortex was chosen because it was a likely site for discriminating proteomic differences in ALS patients with and without dementia, and represents an area in which TDP-43 pathological burden has been mapped in FTD patients (Brettschneider, Del Tredici et al. 2014). ANOVA comparisons of each of the disease groups and controls yielded subsets of significantly altered proteins (**Table 5.1**). Interestingly, the number of significantly altered proteins increased when comparing controls to ALS, ALSFTD, and FTD, which is likely due to the presence of clinical dementia and the associated pathological burden in the frontal cortex (**Figure 5.1A**). For supervised clustering analysis, we selected 165 proteins that had at least 2 ANOVA Tukey pairwise comparisons of high significance ( $p < 0.01$ ) among the 6 possible comparisons among the clinical groups. The over-represented gene ontology (GO) terms within these 165 information rich proteins included cytoplasmic vesicle part, clathrin coat, and plasma membrane (**Table 5.2**). Using these differentially expressed protein signatures of disease, we used multidimensional scaling (MDS) to stratify the relatedness of individual cases within a dataset, which revealed that clinical phenotypes indeed associate with proteomic signatures, as samples segregated into clusters representing clinical groups (**Figure 5.1B**). Also, ALS/FTD cases distributed between the ALS and the FTD clusters, supporting an underlying molecular, biological spectrum designated by protein expression that defines the clinical spectrum. Thus, this analysis demonstrates that

protein signatures of FTD-ALS clinical phenotypes are related to differences in molecular pathways in the frontal cortex.

#### *Protein co-expression network analysis*

Protein co-expression networks reflect relationships between protein pathways, cell types, and physically interacting proteins within modules. Network analyses revealed 15 modules of strongly co-expressed groups of proteins (**Table 5.1**). Each module is defined by an eigenprotein, the most representative weighted protein expression pattern across samples for a group of co-expressed proteins (Seyfried, Dammer et al. 2017). Eight of these 15 modules, identified by numbers that correspond to a color ordered 1 to 15, with M1 (largest, 548 proteins) to M15 (smallest, 21 proteins), showed different patterns of co-expression across the 4 clinical groups. As expected, functional annotation of frontal cortex modules classified the proteome into modules associated with specific gene ontologies and brain cell types among other biological sources of co-expression (Gaiteri, Ding et al. 2014). Of the significant modules, 4 (M2, M6, M9, M10) showed increased expression in FTD cases compared to ALS and control cases (**Figure 5.2A**). These modules were enriched for protein transport, inflammation, and RNA binding proteins (**Table 5.2**). One of the significant modules, M15, enriched for blood microparticles and circulating immunoglobulin complexes, showed increased protein expression in all disease groups compared to controls, consistent with a common mechanism of blood brain barrier breakdown in neurodegenerative diseases (**Figure 5.2A**) (Carvey, Hendey et al. 2009, Seyfried, 2017 #12)). The remaining 3 significant modules (M1, M3, M8) showed a decrease in protein expression in FTD cases compared to control and ALS cases; M1 and M8 were enriched in synaptic and neuronal proteins, while M3 was enriched with mitochondrial proteins (**Figure 5.2A, Table 5.2**). Moreover, several modules showed significant correlation with clinical and

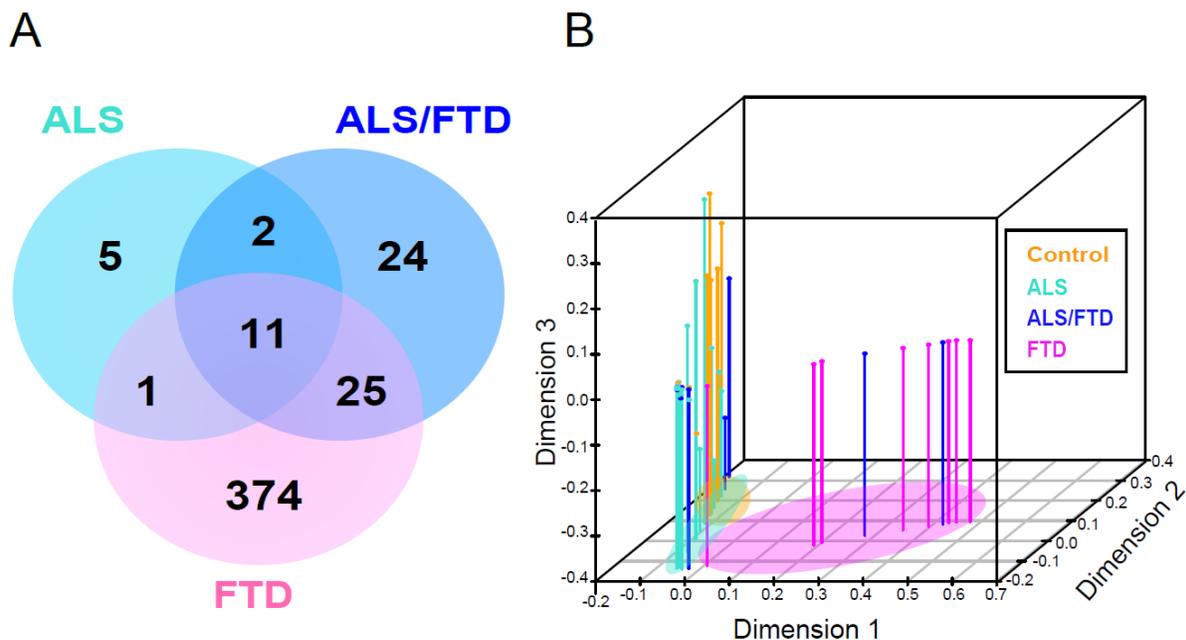
pathological sample traits. Pathological sample traits characterize TDP-43 pathology, which is represented by the abundance of phosphorylated TDP-43 cytoplasmic inclusions in the frontal cortex (defined by the pTDP score) or label-free protein quantification (LFQ) of TDP-43 (**Figure 5.3**). For example, M5, a module enriched with extracellular matrix and astrocyte proteins correlated closely to clinical and pathological traits (**Figure 5.4A**) and M14, a module enriched with microtubule proteins with dynactin as a hub protein, and optineurin, genes previously implicated in ALS and FTD (Fecto and Siddique 2011), correlated with pathological TDP-43 scores and TDP-43 levels. Additionally, M7, enriched with proteins involved in protein transport, correlated with TDP-43 pathology (**Figure 5.4A**).

As expected for the frontal cortex of ALS cases, where TDP-43 pathology is typically not found, and no cognitive decline phenotype exists, module eigenprotein expression for ALS patients without dementia was similar to controls. There was little overlap between controls and either FTD or ALS/FTD. Notably, we found differences in eigenprotein expression between ALS/FTD and FTD alone. Namely, the eigenprotein expression for many of the modules representing the ALS/FTD group was intermediate between that of the strictly ALS and strictly FTD clinical groups, supporting the existence of a molecular spectrum of disease. Mapping of differentially expressed proteins to co-expression modules in cases with ALS/FTD and FTD further supports an underlying molecular continuum (**Figure 5.2B**). There was an enrichment of differentially expressed proteins altered in pairwise comparisons comparing ALS/FTD or FTD patients to controls that were negatively correlated to neuronal modules (M1 and M8) and positively correlated to M2, M6, M9, and M10 modules. Additionally, the proteomic network generated in this study was compared to a previously generated proteomic network created from frontal

cortex samples of Emory ALS, AD, and PD cases (Seyfried, Dammer et al. 2017). Protein co-expression networks were highly preserved when comparing both protein networks (**Figure 5.5A**) and over representation analysis showed that all 15 modules generated in the current study had at least one cognate module within the previously generated Emory brain protein network (**Figure 5.5B**). The consistency between these independent datasets provides confidence in the networks generated in this study, and provides opportunity to expose meaningful relationships along the ALS-FTD disease spectrum by relating clinical and pathological sample traits to modules.

**Table 5.1. Differential and co-expression protein analysis (Supplemental file 3)**

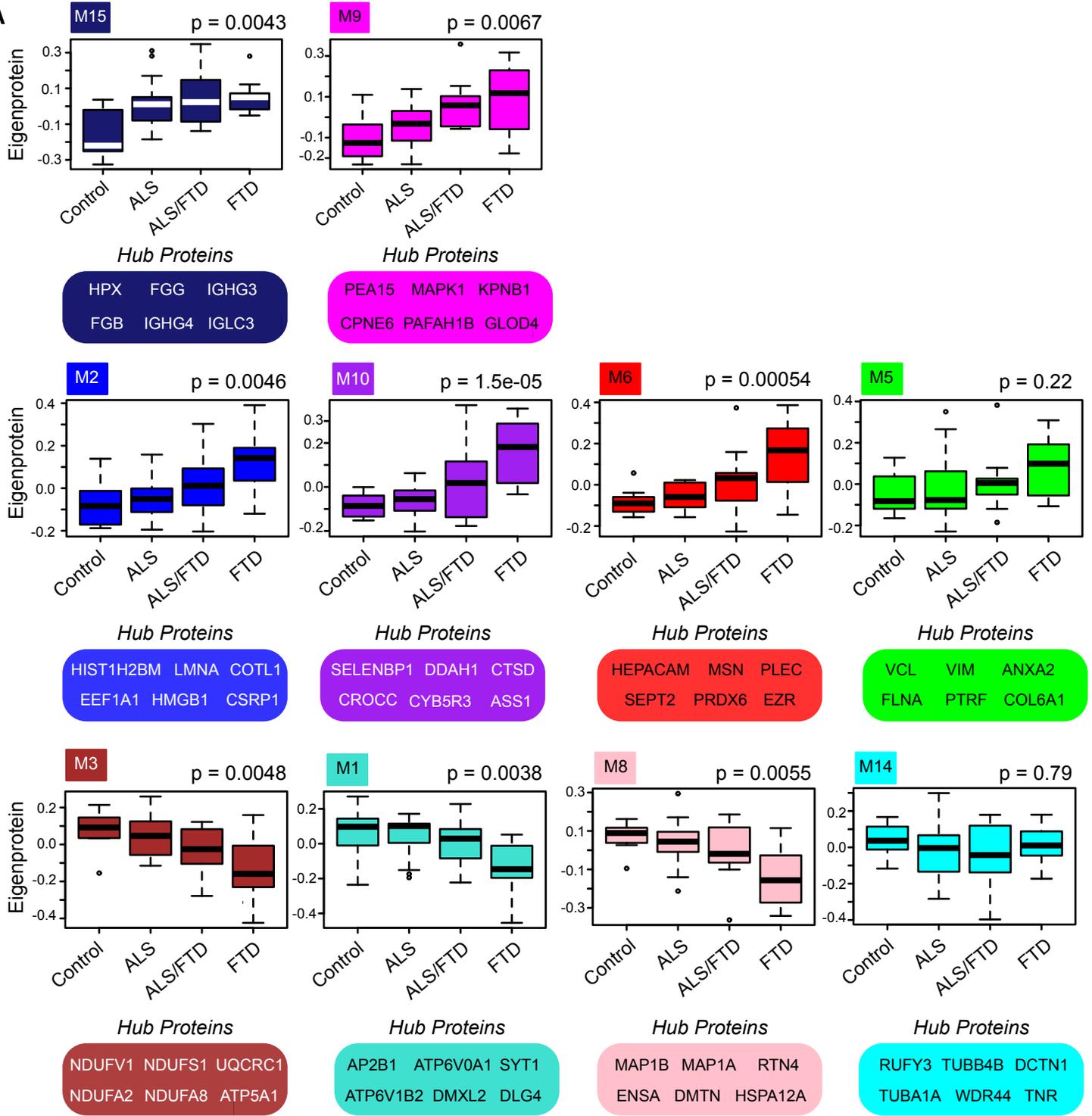
**Table 5.2. Gene ontology (GO) term enrichment for co-expression network (Supplemental file 4)**



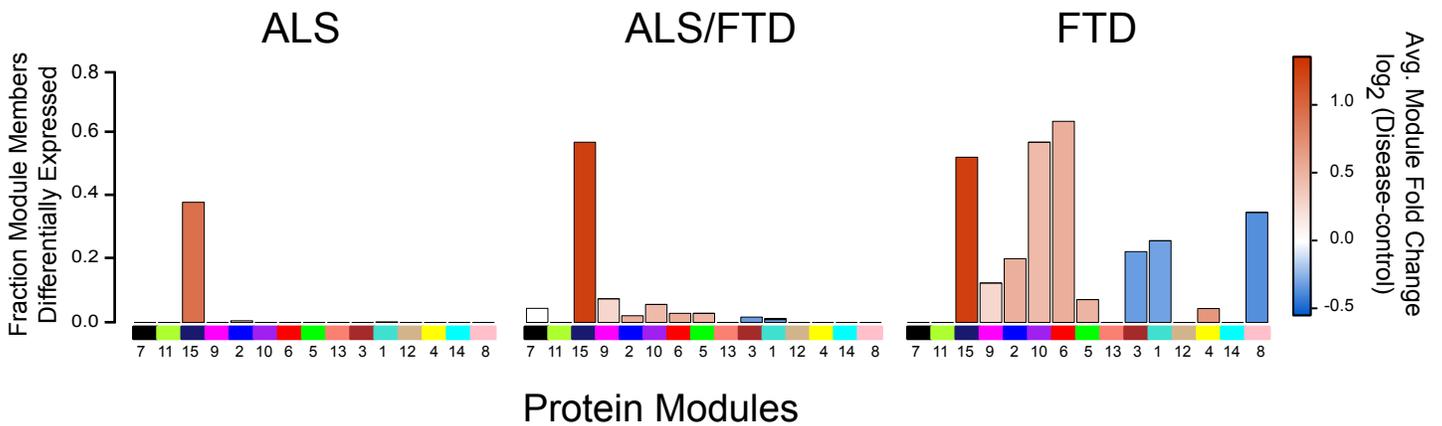
**Figure 5.1. Segregation of clinical ALS-FTD phenotypes by proteomic signatures**

**A.** Venn diagram showing a total of 442 unique proteins determined to be significantly altered by ANOVA followed by Tukey's post-hoc test ( $p \leq 0.01$ ) among 3 pairwise comparisons (i) ALS, (ii.) ALS/FTD, and (iii) FTD versus control cases. **B.** Supervised hierarchical clustering of 165 significant proteins altered in the frontal cortex across clinical phenotypes (CTL, ALS, ALS-FTD, and FTD) by criteria described in methods is shown on by a 3-dimensional multi-dimensional scaling (MDS) analysis. Each dimension on the plot explains a larger proportion of variance of the dataset. Clinical groups are indicated by colors (orange-controls, light blue-ALS, dark blue-FTD/ALS, and pink-FTD). Segregation of cases based on clinical grouping is indicated by colored ovals.

**A**

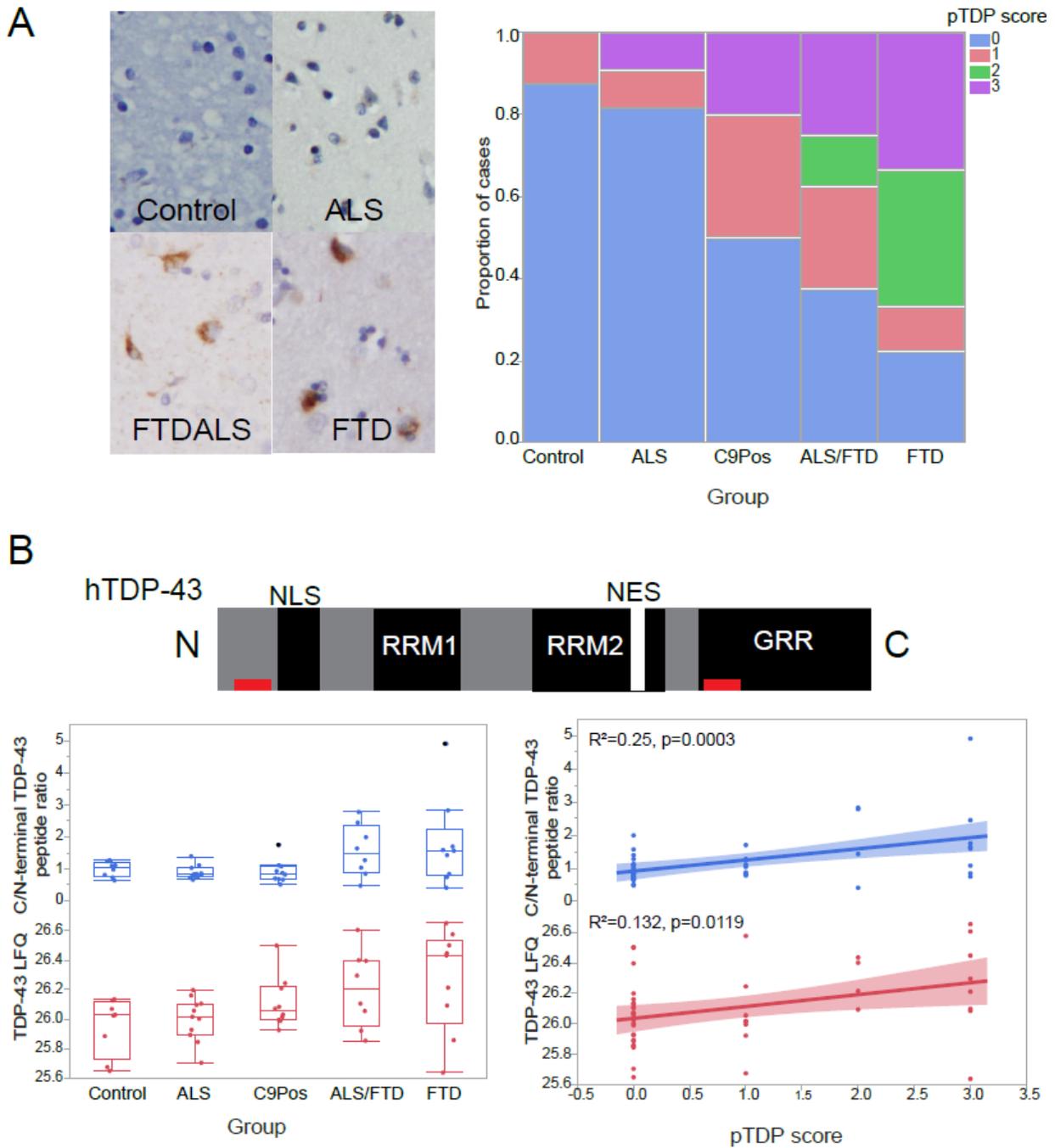


**B**



**Figure 5.2. Integrated protein co-expression and differential expression across the ALS-FTD disease continuum**

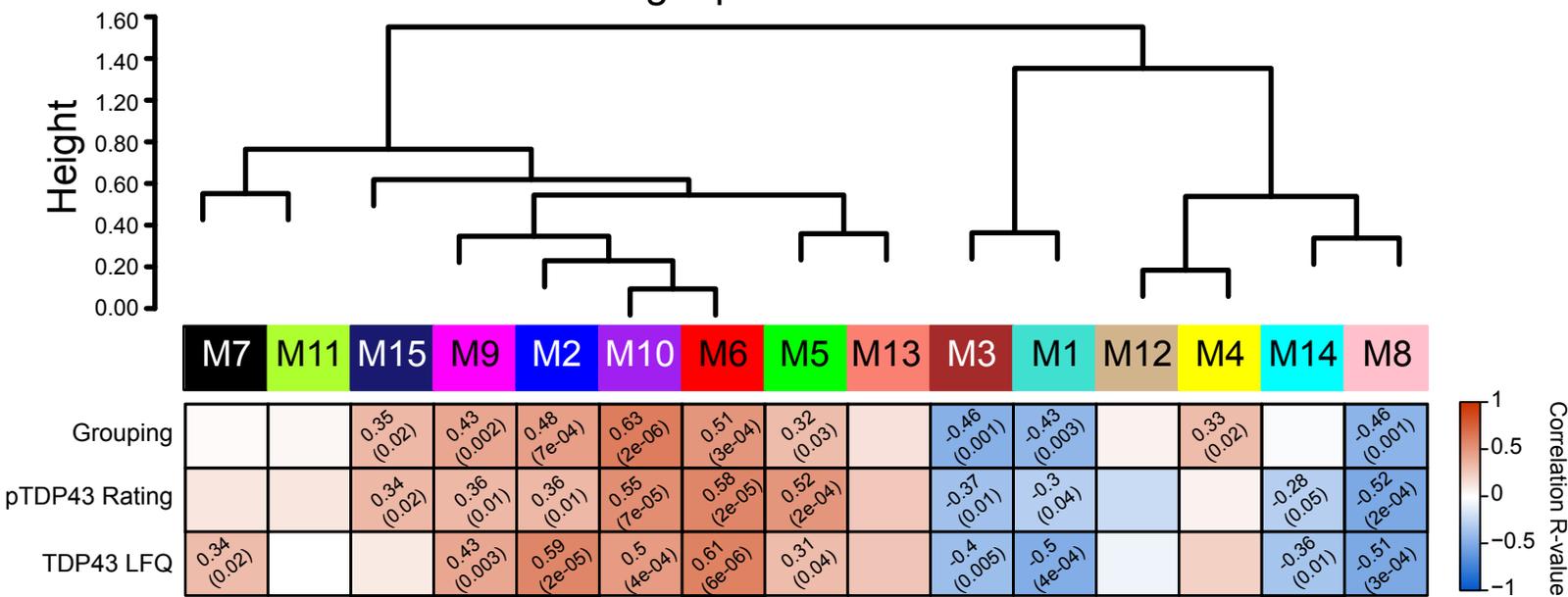
**A.** Module expression box plots are shown (for 10 of the 15 modules) that correspond to the module eigenprotein, which represents the first principal component of a given module and serves as expression profile for the module. Box plots with error bars beyond the 25<sup>th</sup> and 75<sup>th</sup> percentiles are presented for all 4 groups of samples (control, ALS, ALS/FTD, and FTD). Outlier cases are shown as open circles beyond the error bars. Significance was determined by comparisons of quantified proteins within individual patient samples across the clinical groups to the module eigenprotein using one-way nonparametric ANOVA Kruskal Wallis p-values listed above plots. Hub proteins for each of these modules are also highlighted. **B.** Stacked bar plots represent analysis of differential expression (pairwise comparison between listed groups) and enrichment of differentially expressed proteins within co-expression modules. Modules are listed along the x-axis and the height of the bar along the y-axis indicates the proportion of significantly differentially expressed module members, while the color indicates the fold change (red is increased and blue is decreased) according to the scale shown.



pTDP pathological rating score (pTDP score), which is summarized in the mosaic plot to the right. **B.** The human TDP-43 domain structure is shown, including its nuclear localization signal (NLS), RNA recognition motifs, nuclear export signal (NES), and Glycine rich regions (GRR). The C-terminal (FGGNPGGFGNQGGFGNSR; residues 276-293 of human TDP-43, UniProtKB Q13148-1) and N-terminal (LVEGILHAPDAGWG NLVYVVNYPK; residues 56-79 of human TDP-43, UniProtKB Q13148-1) peptides used to generate a C/N-terminal peptide ratio are indicated by red rectangles. Box plots show graphical representations of the C/N-terminal peptide ratio across the groups (blue). Also, shown below are box plots representing total TDP-43 (TDP LFQ) levels across the groups (red). A one-way ANOVA was conducted to compare this ratio across the groups (control (CTL), ALS, FTDALS, and FTD). One-way ANOVA comparing the C-terminal peptide extracted ion chromatogram (XIC), representing abundance, to N-terminal peptide XIC ratio at the  $p \leq 0.05$  level [ $F(3, 43) = 2.6508, p = 0.0607$ ] showed a trend towards increased ratio in FTD cases. There was a significant difference by one-way ANOVA comparing TDP-LFQ levels across the clinical groups at the  $p \leq 0.05$  level [ $F(3, 43) = 3.0530, p = 0.0385$ ]. Correlation plots are shown to the right which display correlation between the C-terminal peptide XIC to N-terminal peptide XIC ratio (blue), TDP-LFQ (red) and pTDP pathological rating; both traits significantly correlated to pTDP score ( $R^2 = 0.251, P = 0.0003$  and  $R^2 = 0.132, P = 0.0119$ , respectively).

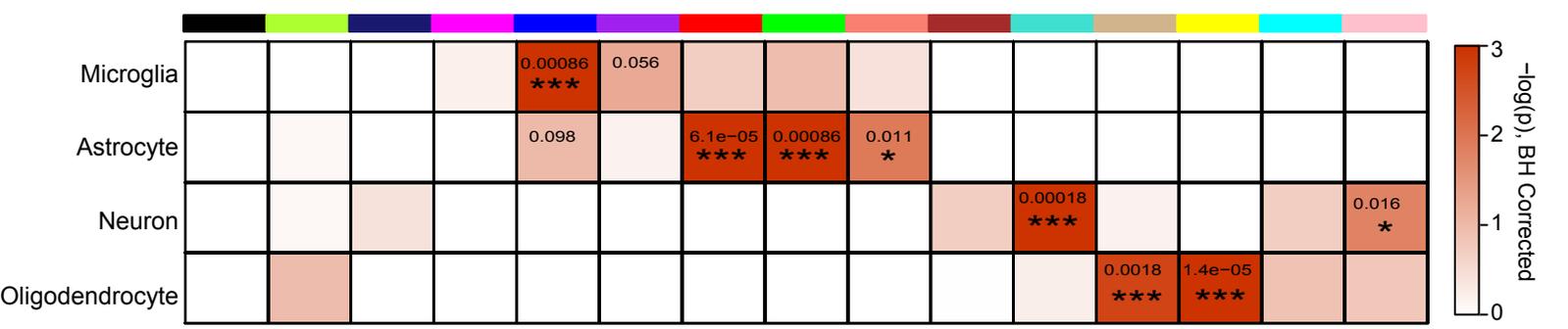
A

### Eigenprotein Network



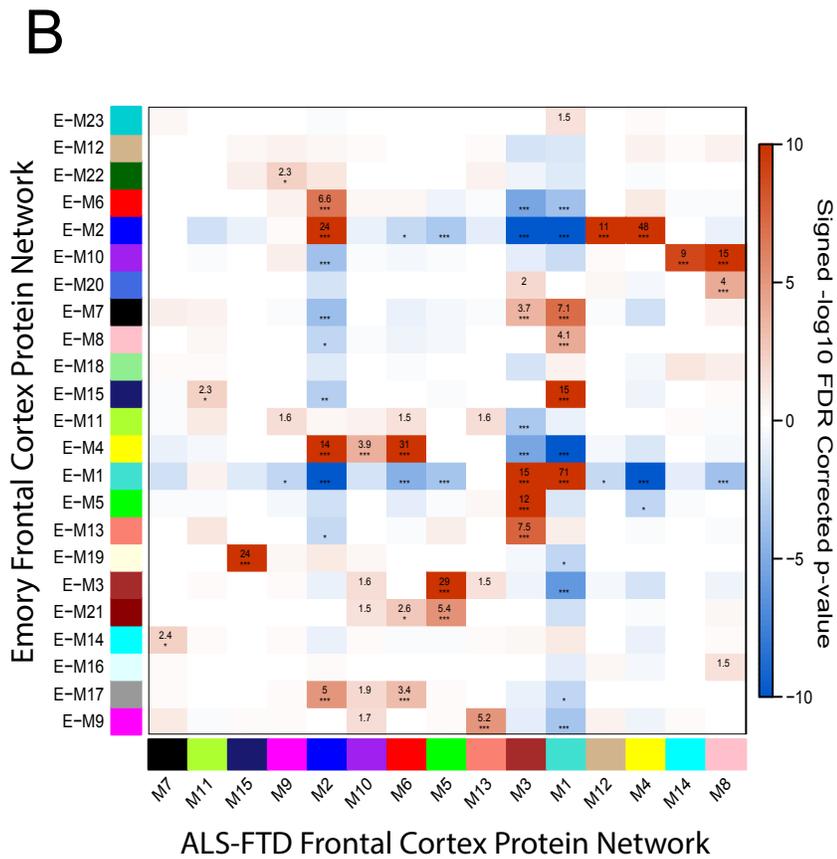
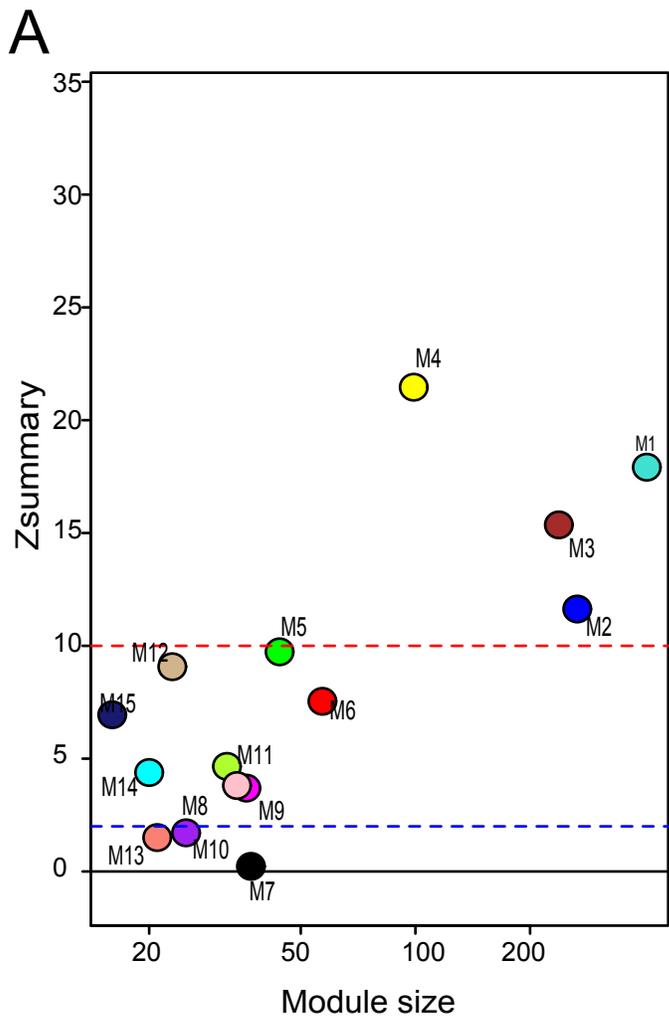
B

### Cell Type Enrichment



**Figure 5.4. Protein co-expression organizes the proteome into modules associated with brain cell types and clinical and pathological traits**

**A.** WGCNA cluster dendrogram groups proteins ( $n=2,613$ ) measured across the frontal cortex into distinct protein modules (M1-M15) defined by dendrogram branch cutting. Modules are clustered based on relatedness defined by correlation of protein co-expression eigenproteins (indicated by position in colorbar). Listed in the heatmap are bicor correlation value and p-value defining relationship between module eigenprotein expression and grouping (clinical grouping defined as 0-control, 1-ALS, 2-ALSFTD, and 3-FTD), pTDP-43 rating (defined as a score from 0-3 that represents pathological TDP-43 burden in the frontal cortex), and TDP-43 label free quantification (LFQ). **B.** Cell-type enrichment analysis was performed using a one-tailed Fisher's exact test against lists of proteins previously generated from acutely isolated neurons, oligodendrocytes, astrocytes and microglia. The heatmap displays Benjamin Hochberg corrected p-values (to control FDR for multiple comparisons) for the enrichment of certain cell types (vertical axis), and protein modules (horizontal axis) indicated by module number and color. Significance is demonstrated by the color scales which go from 0(white) to 3(red), representing  $-\log(p)$ .



**Figure 5.5. Validating the ALS-FTD protein network against previously generated Emory Brain protein network from multiple neurodegenerative diseases**

A. Preservation Z-summary test of protein modules in frontal cortex comparing total homogenate proteome to previously generated Emory Brain Network from label-free proteomic analysis of the frontal cortex from controls, Alzheimer's Disease (AD), Parkinson's Disease (PD) and ALS cases. Less than 0 represents no preservation; 0–2, weak preservation; 2–10, moderate preservation; and more than 10, high preservation. 12 of the 15 modules generated in the total homogenate proteomic network were preserved in the Emory Proteomic network, 8 had moderate preservation with a Z-summary score between 2 ( $p < 0.05$ ) (blue dotted line) and 10 ( $p < 0.01$ ) (red dotted line), while 4 were highly preserved with a Z-summary score greater than 10 ( $p < 0.01$ ). **B.** Over-representation analysis for protein networks generated from frontal cortex tissue (Emory Brain Proteomes) using a two-sided Fisher exact test with 95% confidence intervals. P-values were FDR adjusted to account for multiple comparisons. The 15 modules in the total homogenate network were aligned to the 23 modules in the Emory case network. Colors represent whether module gene symbol lists showed significant overlap (red), depletion (blue) or no significant under/over representation (white) for protein membership. Numbers displayed on the heatmaps represent positive signed  $-\log_{10}$ (FDR-adjusted p values), asterisks represent the level of significance of comparisons (\* $p < 0.05$ ; \*\*,  $p < 0.01$ , \*\*\*,  $p < 0.005$ ).

*Protein co-expression classifies the proteome into modules associated with brain cell types*

Co-expression organizes the proteome by placing proteins as nodes within a network with edges that define the connectivity, correlation, of nodes to one another (Gibbs, Baratt et al. 2013). Many biological networks have an inherent hierarchical structure that allows nodes to be organized into a small number of highly interconnected modules (Parikshak, Gandal et al. 2015). Intra-modular connectivity creates a rank for proteins within a module, which allows identification of proteins that are module hubs; these are enriched for potential key drivers of a module's co-expression (Parikshak, Gandal et al. 2015). Several modules were enriched for microglial & astrocytic proteins (M2, M10, M6, M5, M13), while others were enriched for neuronal proteins (M1 & M8), or oligodendrocyte specific proteins (M12 & M4) (**Figure 5.4B**), consistent with previous studies (Miller, Woltjer et al. 2013(Oldham, 2008 #379) (Seyfried, Dammer et al. 2017)). Modules of co-expressed proteins enrich with markers of specific brain cell types, and so changes in protein co-expression that relate to clinical phenotypes along the ALS-FTD disease spectrum may reflect changes in the phenotypes, abundance, and/or activation of specific cell types (Seyfried, Dammer et al. 2017). Module gene ontology enrichment also typically closely mirrors the cell-type enrichment analysis results (**Table 5.2**). For example, the M6 module was enriched for proteins involved in responding to biotic stimuli consistent with the cell-type enrichment analysis of astrocytic and microglial proteins with known roles in neuroinflammation. Modules enriched for gene ontologies associated with “synapse-part (GO:0044456)” proteins and “neuronal differentiation (GO:0030182)”, M1 and M8, respectively, are enriched for neuronal-specific proteins. Microglial and astrocytic modules (M2, M6, M5, M13) were positively correlated to pathological TDP-43 rating and clinical grouping, while neuronal modules (M1, M8) were negatively correlated to these traits (**Figure 5.4A**). M4,

enriched with oligodendrocyte proteins, also correlated moderately to ALS-FTD clinical grouping (**Figure 5.4A**). Together, this suggests cell-type specific processes undergo changes in ALS/FTD that manifest as clinical and pathological traits, some of which are similar to changes occurring with AD (De Strooper and Karran 2016, Seyfried, Dammer et al. 2017).

#### *Module relevance to FTD pathology and clinical grouping*

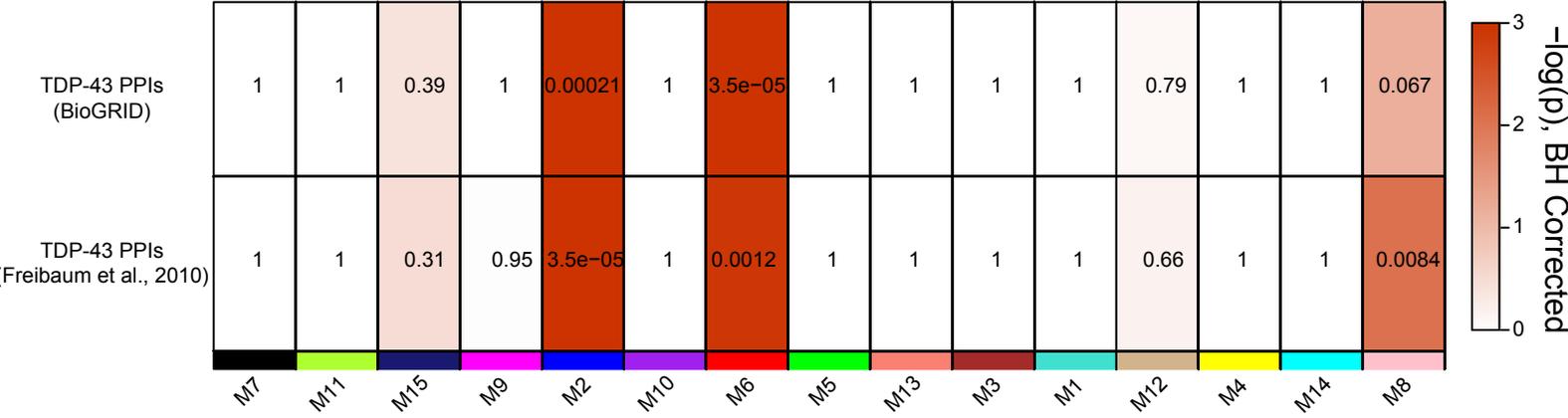
The M2 and M6 modules, enriched for RNA binding proteins and inflammatory proteins, respectively, were strongly positively correlated to TDP-43 pathological burden, total TDP-43 levels and clinical grouping. TDP-43 mapped to M2, a module of co-expressed proteins increased significantly in ALS/FTD and FTD compared to controls. Notably, this module also included significant enrichment of microglial proteins (hypergeometric overlap  $p < 0.001$ ), suggesting a strong co-expression between RNA binding protein dysfunction and microglial inflammation in the ALS-FTD spectrum that was not apparent for other neurodegenerative cohorts (Seyfried, Dammer et al. 2017). As expected, immunohistochemical analysis of pTDP pathology showed a significant increase in FTD cases compared to ALS and control cases, which mirrors protein co-expression data involving overall TDP-43 protein abundance in the frontal cortex of these brains (**Figure 5.3**). Previous work has suggested that the C-terminal peptide accumulation of TDP-43 is associated with potential toxicity of pathological TDP-43 (Zhang, Xu et al. 2009). Indeed, we observed an increase in C-terminal TDP-43 when analyzing peptide level differences (**Figure 5.3**). Thus, the correlation of inflammatory proteins and RNA binding proteins in the M6 and M2 modules, respectively, to clinical (M6) and pathological (M2) traits, may represent a molecular pathophysiological connection between TDP-43 proteinopathy,

related post-translational modifications, and consequent changes in brain proteome occurring along the ALS/FTD clinical disease spectrum.

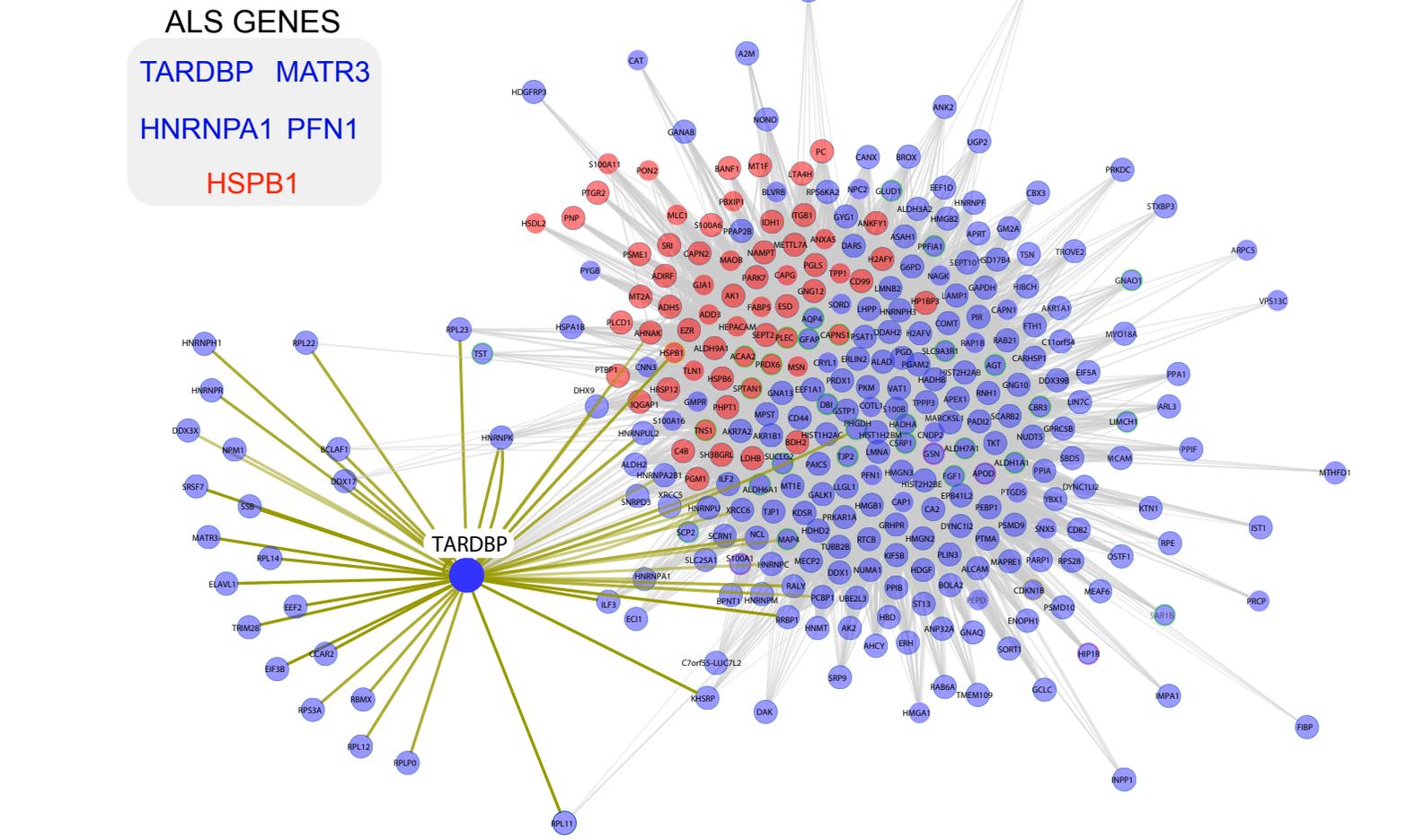
*Modules associated with inflammation are enriched with TDP-43 interactors and have causal links to ALS*

Modules associated with TDP-43 pathology and clinical phenotypes described above could either play a causal role in FTD or be secondary to the disease process. While postmortem protein expression does not directly assess causality of these modules by themselves, integrating multiple ‘omic’ data sources can prioritize those modules that are most central to FTD and ALS pathogenesis. One such approach for module prioritization is to assess the enrichment of known TDP-43 protein-protein interactions (PPIs) within each respective module generated for co-expression across the ALS-FTD spectrum of cases. For example, the M2 and M6 modules were significantly enriched for proteins previously identified in studies exploring TDP-43 PPIs (**Figure 5.6A**). Furthermore, the protein products for several genes that have been causally linked to ALS (Taylor, Brown et al. 2016) are found within the M2 module, including several hnRNPs (HNRNPA1), MATR3, and PFN1 and TDP-43 itself (**Figure 5.6B**), whereas HSPB1 (in M6) has been linked to hereditary motor neuropathy, a form of motor neuron disease (James and Talbot 2006, Rossor, Kalmar et al. 2011). Additionally, the M8 module, enriched with synaptic proteins, showed enrichment with TDP-43 PPIs, yet lacked any genes causally linked to ALS or FTD. The strong enrichment of TDP-43 interactors and other causal genes for ALS within M2 and M6 modules further reinforces their association with the ALS/FTD spectrum and suggest that proteins within these modules have critical roles in mechanisms that drive TDP-43 aggregation and other cellular driven pathological processes (i.e., neuroinflammation).

**A**



**B**



**Figure 5.6. Modules with causal links to ALS are associated with inflammation and enriched with TDP-43 protein-protein interactions (PPIs)**

**A.** Enrichment of TDP-43 PPIs is displayed in this heatmap with results from TDP-43 PPI from BIOGRID and those from published global analysis of TDP-43 interacting proteins, with protein co-expression modules listed across the horizontal axis. Colors on heat map indicate enrichment (red) or no significant over-representation (white) for gene membership. Numbers displayed on the heatmaps represent positive signed  $-\log_{10}(\text{BH-adjusted } p \text{ values})$ , asterisks represent the level of significance of comparisons (\* $p < 0.05$ ; \*\*,  $p < 0.01$ , \*\*\*,  $p < 0.005$ ). **B.** M2 and M6 modules represented by proteins with high co-expression (white lines) and TDP-43 PPIs (yellow lines). Several genes that have been previously linked to ALS-FTD are highlighted as yellow circles.

*Astrocyte and microglial markers associated with neuroinflammation are increased across the ALS-FTD spectrum*

The M6 module was significantly enriched for TDP-43 PPIs, and positively correlated to pathological TDP burden and clinical dementia. This module includes hepatic and glial cell adhesion molecule (HEPACAM), membrane-organizing extension spike protein moesin (MSN), ezrin (EZR), glial fibrillary acidic protein (GFAP) and peroxiredoxin 6 (PRDX6). HEPACAM, an astrocyte specific protein (Zhang, Chen et al. 2014), is involved in regulating cell motility and cell-matrix interactions, while moesin and ezrin are members of a family of proteins which function as cross linkers between plasma membranes and actin-based cytoskeleton, also regulating motility (Ivetic and Ridley 2004). Many of the proteins in the M6 module are expressed in glial cell types (astrocytes and microglia), and have gene ontologies associated “response to biotic stimuli” thus defining what we characterize as the inflammatory module (**Figure 5.7A**). Module eigenprotein expression revealed M6 as a module increased in FTD compared to controls and ALS. The WGCNA measure of intramodular connectivity ( $k_{ME}$ ) defined as the Pearson correlation between the expression pattern of a protein and the module eigenprotein (which summarizes the characteristic expression pattern of proteins within a module), quantifies the extent to which individual proteins track this pattern (Seidel, Marangoni et al. 2017). High relative  $k_{ME}$  can be used to identify individual proteins that best represent a module; typically, these hubs of modules are markers of predominant cell types within those regions (Parikshak, Gandal et al. 2015). Four proteins, HEPACAM, MSN, GFAP, and TPP1, with high  $k_{ME}$  (i.e., hub-like module membership), were confirmed to be increased in abundance in frontal cortex homogenates from FTD cases compared to that of control and ALS cases by Western blot analysis (**Figure 5.7B**). The differences that define co-expression can also be

appreciated in these Western blots as the respective protein levels for each target in the samples change in tandem, representing correlated co-expression. A subset of cases from each clinical group was also immunostained with antibodies against MSN and GFAP to respectively represent microglial as well as astrocytic cellular populations, and increased protein expression in FTD cases compared to controls was also clear (**Figure 5.7C**).

Protein co-expression analysis resolved expected biological differences in the frontal cortex in FTD, including reduced neuronal or synaptic populations, as represented by reduced eigenprotein expression in FTD and ALS/FTD cases specific for modules enriched with neuronal and synaptic proteins (M1 and M8) (Kersaitis, Halliday et al. 2004). The eigenprotein for M1, a module enriched for synaptic signaling proteins, was significantly reduced in FTD cases compared to controls and ALS. M6 and M1 protein co-expression and status-dependent changes demonstrates the robustness of our integrated systems approach and highlights important biology underlying potential mechanisms of disease. We found increased inflammation within the frontal cortex in FTD cases, but not ALS, which correlates with TDP-43 abundance, suggesting a potential role of immune cells in disease pathogenesis and progression related to TDP proteinopathy.

#### *Co-expression defines C9orf72 specific changes in ALS brain related to neuroinflammation*

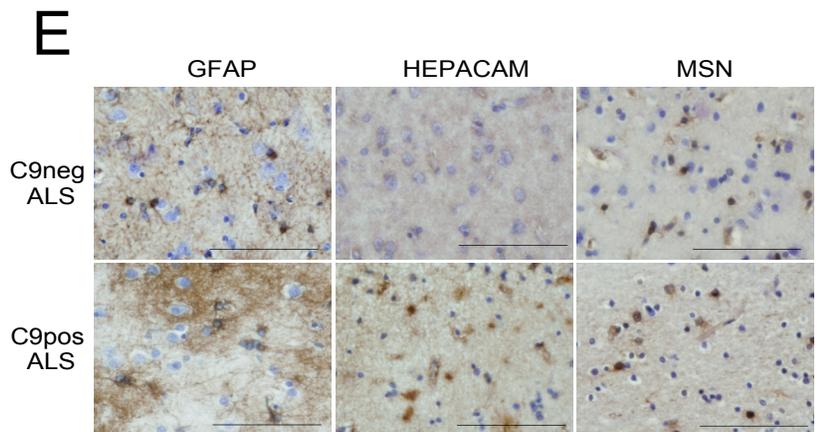
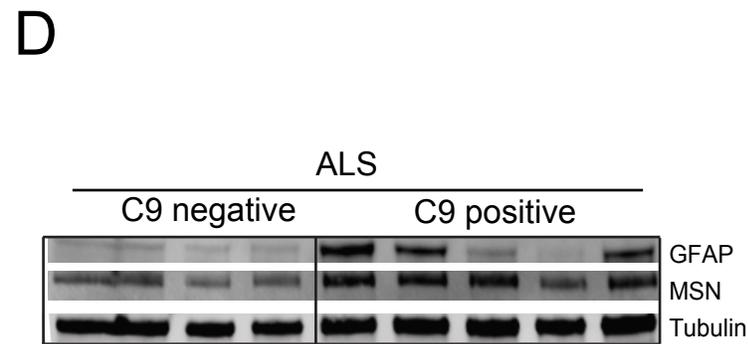
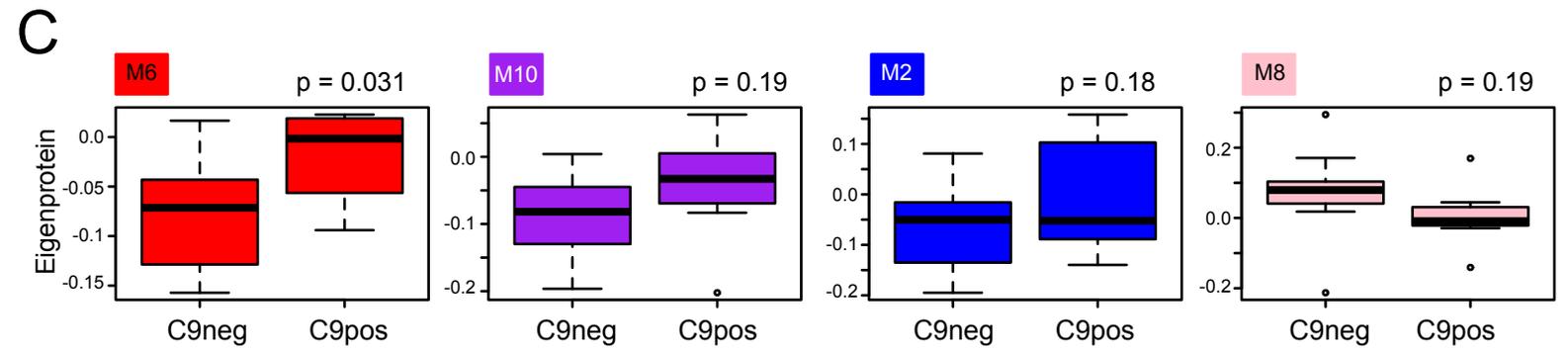
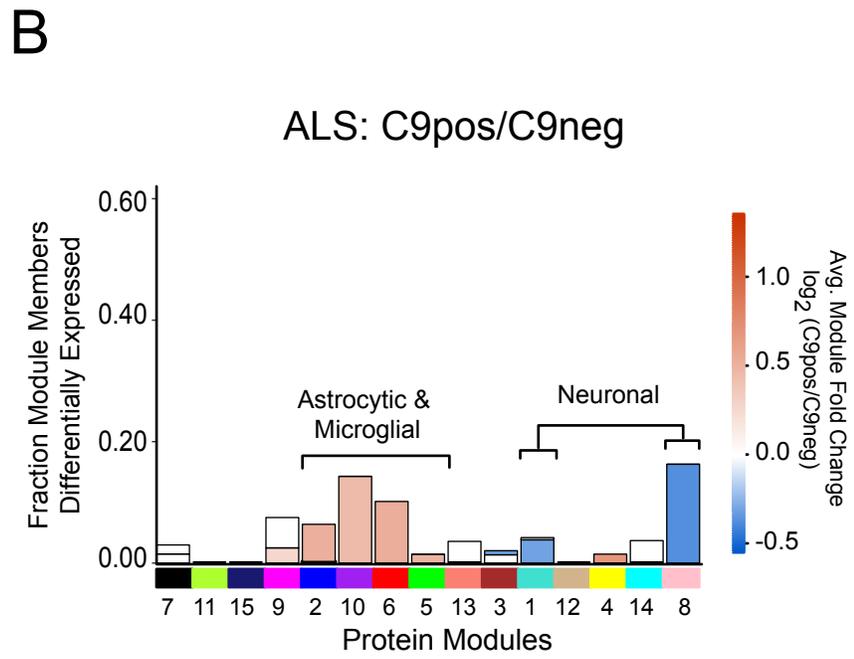
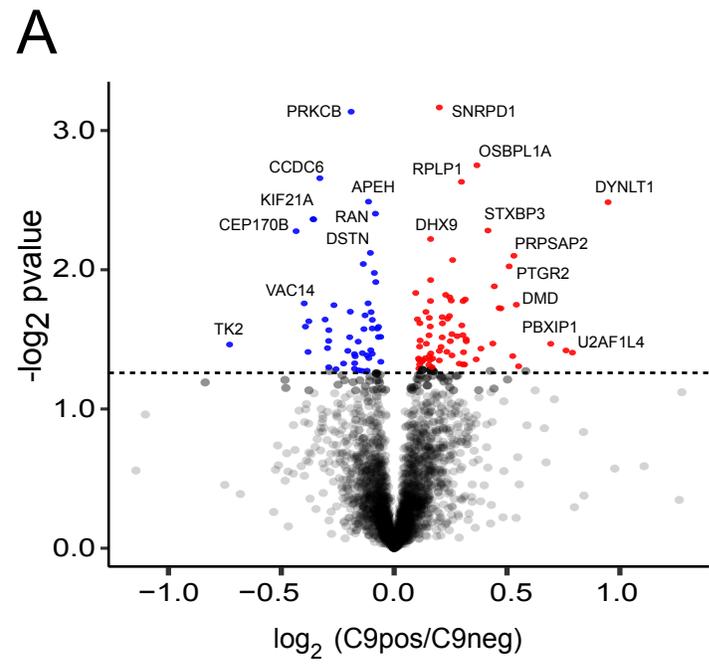
To elucidate the contribution of the C9orf72 genetic expansion to disease we compared protein expression relative to our C9Pos cohort. Few proteins were differentially expressed in the frontal cortex of C9Pos ALS cases compared to C9Neg ALS cases (**Table 5.1, Figure 5.8A**). Furthermore, none of these C9orf72 genotype-affected proteins were changed by greater than 2-fold, making it difficult to resolve proteome wide differences by differential expression alone.

The presence of the C9orf72 expansion as a case-sample trait also did not significantly correlate with any of the identified protein co-expression module eigengenes in the network when comparisons were drawn across all clinical groups. However, mapping of the differentially expressed proteins in C9Pos ALS to the same modules from the co-expression network revealed that those proteins with significant, yet marginal, fold change in C9Pos ALS cases, generally mapped into the astrocytic/microglial modules (positively correlated with gene expansion) and neuronal modules (negatively correlated), consistent with changes associated with cognitive dysfunction in FTD (**Figure 5.2B**). Furthermore, in an analysis focused on strictly comparing only the non-demented ALS patients with and without the C9orf72 expansion, we identified one module eigenprotein (M6) ( $p < 0.05$ ) as increased in C9Pos ALS (**Figure 5.8C**). These patients were not demented prior to death, suggesting that the C9orf72 mutation may be associated with neuroinflammation in the brain. These results also highlight the ability of our integrative systems approach with WGCNA to resolve genetic vulnerability, which could not be confidently recognized through differential expression analysis alone. The proteomic findings of increased inflammation were validated by showing relative increased expression of three (MSN, GFAP, and HEPACAM) microglial, astrocytic, and inflammatory proteins in the frontal cortex of C9Pos ALS patients by Western blot (**Figure 5.8D**) and immunohistochemical analyses (**Figure 5.8E**). Although not significant ( $p = 0.13$ ), we found a relative reduction of the C9orf72 protein (~55kDa long isoform) in C9Pos ALS cases compared to C9Neg ALS cases consistent with a loss of C9orf72 function (**Figure 5.9**). This is of interest because haploinsufficiency of C9orf72 protein, possibly related to neuroinflammation, has been suggested as a possible disease mechanism in C9Pos patients (Farg, Sundaramoorthy et al. 2014, Waite, Baumer et al. 2014, Yang, Liang et al. 2016).



**Figure 5.7. Astrocyte and microglial markers associated with neuroinflammation are increased across the ALS-FTD spectrum**

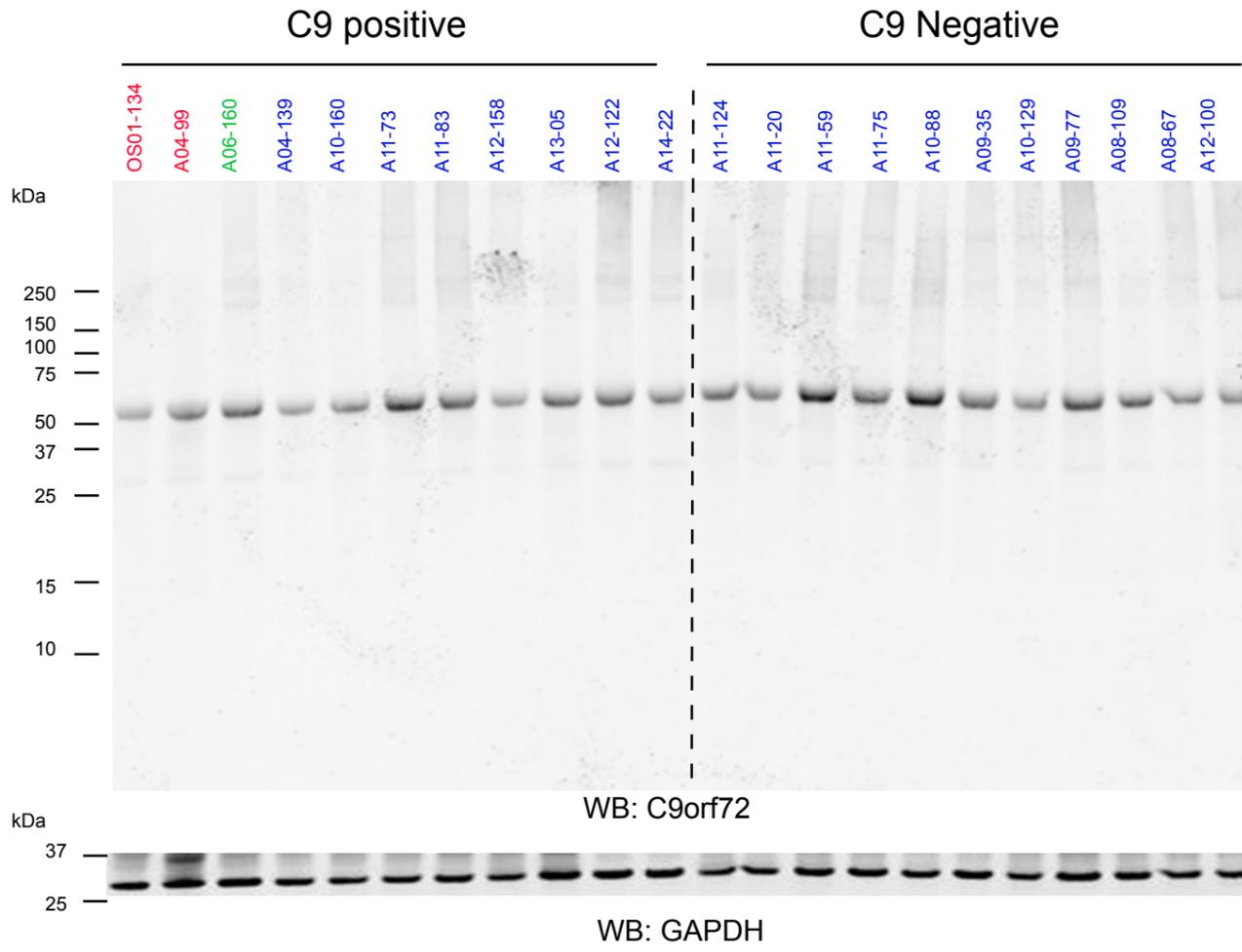
**A.** I-graph, representing proteins within the M6 (red) co-expression module, presents genes with high module membership more centrally and cell type is demonstrated by rim and text color around the gene symbols (blue: microglial, and green: astrocytic). The M6 eigenprotein showed increased expression in FTD and ALS-FTD cases compared to CTL and ALS cases. **B.** Western blot results across individual cases in each of the groups are shown and quantified in the graph below (bars represent means and error bars the SEM); these display quantitative difference in relative expression of black module proteins (p-values for respective ANOVAs were: 0.0018 for GFAP, 0.0117 for HEPACAM, 0.0014 for TPP1, 0.2476 for MSN). **C.** Immunohistochemical validation visually showed a difference in the M6 module protein expression in astrocytes and microglia across the groups. Representative images at 40X magnification for an ALS (top) and ALS/FTD (bottom) case are shown for GFAP and HEPACAM, astrocytic staining in cells of the cortical layers, demonstrating an increase in astrocytes in the frontal cortex in FTD cases compared to ALS. MSN staining demonstrates increased microglial staining patterns in FTD cases compared to ALS and CTL, consistent with Iba-1 (microglial marker) immunoreactivity. Scale bars represent 200 $\mu$ m.



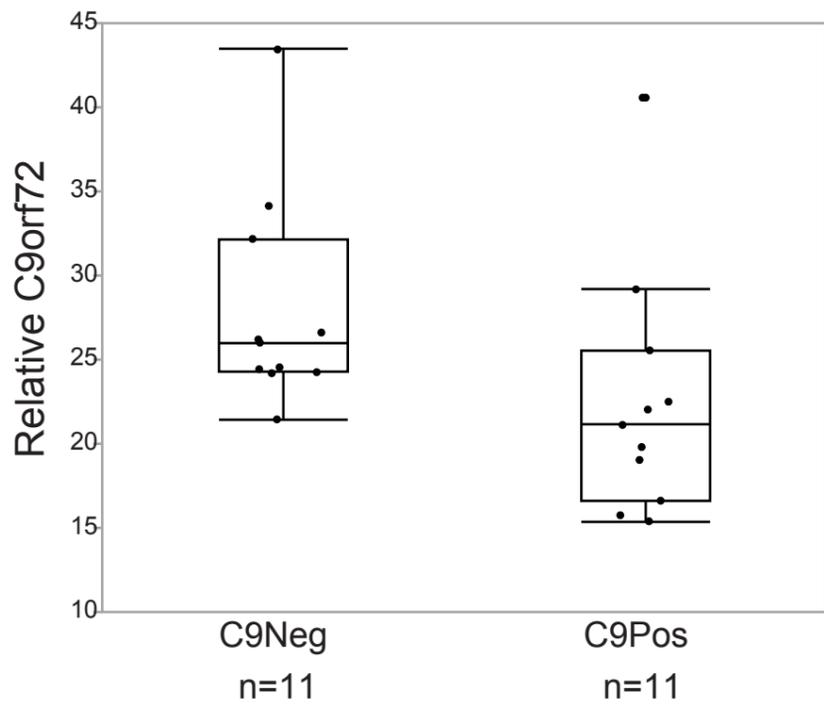
**Figure 5.8. Co-expression analysis resolves C9orf72 specific changes in ALS brain**

**A.** Volcano plot displays the protein abundance ( $\log_2$  fold change) against the t-statistic ( $-\log_{10}(\text{p-value})$ ) for C9Pos (n=8) against C9Neg (n=11) ALS cases. Red and blue dots indicate significantly altered proteins that are increased and decreased respectively in C9ORF72 cases. Gray dots below the hashed line represent proteins that are not significantly changed ( $p > 0.05$ ). Notably, there are no significant proteins that are 2-fold or more ( $\pm \log_2 1.0$ ) increased or decreased in C9Pos cases compared to C9neg cases. **B.** Stacked bar plot represents the analysis of differential expression (pairwise comparison between C9Pos and C9Neg ALS cases) and enrichment of differentially expressed proteins within co-expression modules. Modules are listed along the x-axis and the height of the bar along the y-axis indicates the proportion of significantly differentially expressed module members, while the color indicates the fold change according to the scale shown. **C.** Module expression box plots for 4 of the 15 modules generated comparing C9Pos ALS and C9Neg ALS cases are shown that correspond to the module eigenprotein (the overall protein expression profile of a module). The M6 module is significant, whereas other glial modules (M10 and M2) and a synaptic module (M8) are increased and decreased in C9Pos cases, respectively, but don't reach significance using a Student's t-test (p-value listed above the plot). **D.** Western blot analysis from a subset of cases demonstrated increased expression of M6 module proteins (MSN and GFAP) in C9Pos ALS cases compared to C9Neg cases. Tubulin was used as a loading control. **E.** Immunohistochemical validation showing representative images at 40X from frontal cortex of C9Pos ALS and C9Neg ALS cases for proteins within the black module, GFAP, HEPACAM, MSN, and Iba-1, a microglial marker. Scalebars represent 200 $\mu\text{m}$ .

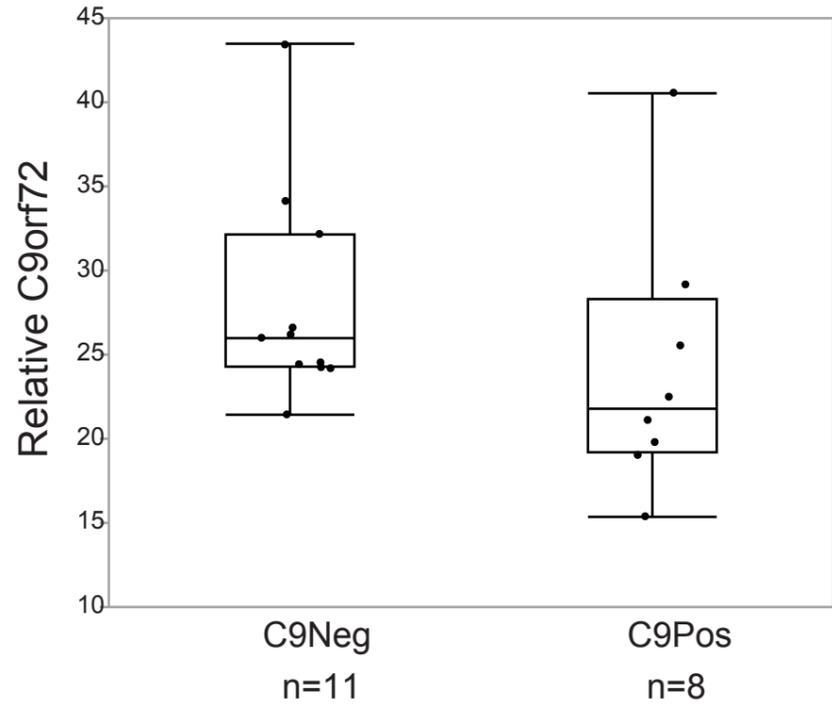
■ FTD ■ FTD/ALS ■ ALS



C9Neg ALS vs C9Pos (n=22)  
p=0.04



C9Neg ALS vs C9Pos ALS (n=19)  
p=0.13



**Figure 5.9. C9orf72 protein reduction in C9Pos ALS compared to C9Neg ALS**

Western blot analysis probing C9orf72 protein (~55 kDA) across individual C9Pos cases and C9Neg ALS cases. There is a decrease in C9orf72 protein expression in C9pos cases demonstrated by quantitative difference in protein expression, relative to GAPDH protein loading control (t-test p-value=0.04 comparing 11 C9Neg ALS cases to 11 C9Pos cases, T-test p-value=0.13 comparing 11 C9Neg ALS cases to 8 C9Pos ALS cases).

## Discussion

Using an unbiased proteomic screen of the post-mortem frontal cortex, we were able to identify and quantify protein differences along the ALS-FTD disease spectrum. Employing both differential and co-expression analysis, we demonstrate that protein expression within the brain has a signature that defines the molecular and genetic underpinnings of the clinical and pathological ALS-FTD disease spectrum. WGCNA resolved related protein co-expression patterns or modules representing pathways and brain cell types, and showed a decrease in expression levels for modules associated with neurons and an increased in expression of astroglial and microglial modules associated with cognitive dysfunction and TDP-43 pathology in brain. Strikingly, C9orf72 expansion in patients dying with ALS, but not FTD, predisposed the frontal cortex to elevated co-expressed markers of neuroinflammation, the same ones co-expressed in a concerted increased pattern across FTD and ALS/FTD patients. Thus, our findings in the context of genetic underpinnings of ALS/FTD, and prior protein co-expression networks for AD (Seyfried, Dammer et al. 2017) indicate that neuroinflammation is a common signature of dementia (due to either FTD or AD), while broad RNA-binding protein co-accumulation is more pronounced in the current network organized around frontal cortex of cases on the ALS-FTD clinical spectrum.

The protein co-expression network generated in this study was consistent with those previously reported in AD (Seyfried, Dammer et al. 2017), which revealed biologically relevant modules linked to specific cells types and organelles (e.g. mitochondria). Although the genetic and pathological drivers in the ALS-FTD spectrum are distinct from AD, we observed a consistent downregulation of modules associated with neurons and synapses and upregulation of glial

(microglial and astroglial) modules with increased TDP-43 pathology and cognitive dysfunction in brain. This suggests signatures reflecting relative changes in cellular phenotypes and or abundance are shared across neurodegenerative diseases (De Strooper and Karran 2016). However, a clear distinction between the AD network and our current ALS-FTD network was the definition of the M2 module, a module enriched with RNA-binding proteins and microglial markers. This module showed a strong correlation with clinical and pathological traits of cases on the ALS/FTD spectrum, co-expressed with protein products of genes with causative links to ALS (hnRNPs, matrin 3, profilin 1, and TDP-43), and had a significant enrichment for TDP-43 PPIs. The strong enrichment of TDP-43 interactors and other causal genes for ALS within this module further supports its strong association with the ALS/FTD spectrum and implicates other novel members of this module as having critical roles in TDP-43 biology that potentially influence TDP-43 aggregation or other pathological processes (i.e. microglial-directed neuroinflammation) inherent to ALS and FTD etiology. The convergence of a large number of RNA binding proteins and microglial markers to the same module (M2) is unique to the ALS/FTD clinical disease spectrum, as we did not observe this same pattern of co-expression in the AD network we previously generated (Seyfried, Dammer et al. 2017). Thus, the alteration of RNA-binding proteins in this disease spectrum is concurrent with an increase in microglial-driven inflammation, and cognitive decline unique to the FTD subgroup of our selected cases in the ALS-FTD spectrum.

Recent findings have also demonstrated links between the C9orf72 expansion mutation and inflammatory pathways in C9orf72 animal models (Burberry, Suzuki et al. 2016, O'Rourke, Bogdanik et al. 2016, Sudria-Lopez, Koppers et al. 2016). Evidence also exists that C9orf72

mutation carriers have an increased prevalence of certain autoimmune disorders (Miller, Sturm et al. 2016) increased microglial pathology (Brettschneider, Toledo et al. 2012), and increased thinning in frontal and temporal lobes in neuroimaging studies, compared to sporadic ALS and controls (Floeter, Bageac et al. 2016). Increased abundance of proteins identified within the inflammatory co-expression module in the frontal cortex of ALS/FTD, FTD and C9Pos ALS cases may explain the clinical link between cognitive decline and the C9orf72 expansion (Irwin, McMillan et al. 2013), (Umoh, Fournier et al. 2016). These data, from an unbiased proteomic approach, support the hypothesis that the dementia phenotype represents changes within the brain that correlate with, and are potentially causally linked to, increased neuroinflammation (Bettcher and Kramer 2013). These findings suggest that inflammatory processes play a role in dementia network concerted changes in the frontal cortex, and that further investigation of the networks we describe in this study may provide insights into underlying biological mechanisms driving the overlap of these clinically disparate diseases.

There were several limitations in this work. One is the coverage of the proteome that we were able to achieve with our single-shot LC-MS/MS approach. The number of proteins identified was similar to proteomic coverage seen in other studies using similar approaches (Bi, Li et al. 2017, Seyfried, Dammer et al. 2017). However, this is much lower than studies investigating the transcriptome. Nevertheless, the co-expression patterns measured were robust and reproducible. Another limitation was the relatively small number of FTD cases carrying the C9orf72 expansion; this diminished our power to discern any differences we may have seen due to this genetic mutation across all the clinical groups. Future studies that include additional C9orf72 expansion positive cases from the ALS/FTD and pure FTD will be critical at resolving genetic

drivers of disease. Of interest, though is our finding of increased microglia- and astrocyte-specific proteins in the frontal cortex of C9Pos ALS patients with no clinically detectable dementia.

Ultimately, our results demonstrate the utility of a systems biology approach in understanding complex diseases with underlying co-morbidity. We were able to relate a proteomic signature of disease (i.e. proteotype) to clinical phenotypes and C9orf72 genotype, which has not previously been done. Proteins within co-expression modules could serve as potential biomarkers of disease mechanism or targets to assess therapeutic interventions across the ALS-FTD disease spectrum, and the network itself serves as a scaffold for correlation of additional proteotype features such as concerted changes in disease-specific protein post-translational modifications. Our study confirms that several pathways in addition to brain inflammation are implicated along the ALS-FTD disease spectrum. This network also provides a resource that moves towards a broad and comprehensive understanding of pathway regulation that manifests in co-expression networks with distinguishing characteristics along the ALS-FTD spectrum.

**CHAPTER 6: Summary and future directions**

## Summary of Work

The goal of the studies presented in this dissertation was to uncover mechanisms involved in the overlap of ALS and FTD, two complex neurodegenerative diseases with clinical, genetic, and pathological overlap. Through analysis of clinical, pathological, and proteomic differences in cases along this disease spectrum these studies add to our understanding of the phenotypic presentation, pathological manifestation and molecular pathways involved in genetic and sporadic forms of these diseases. Since identification of the C9orf72 expansion in 2011 the field of ALS and FTD research has advanced at a rapid pace. During this period of enormous progress, our studies on the clinical, pathological, and proteomic differences in patients along this disease spectrum contribute an additional perspective. We provide evidence of an increased inflammatory profile in FTD cases compared to ALS cases along the disease continuum and identify similar increases in C9Pos ALS cases compared to C9Neg ALS cases further connecting these overlapping diseases.

We leveraged the large clinical population at the Emory ALS center to better understand the clinical presentation of ALS in patients with the C9orf72 mutation, as compared to sporadic ALS patients. We found that other than the presence of a family history of ALS and increased prevalence of FTD, these two groups were clinically similar at presentation to the clinic. The distribution of the site of onset of disease in our cohort of C9Pos patients closely mirrored that of C9Neg patients and was typical of other ALS cohorts.(Williams, Fitzhenry et al. 2013) The racial makeup of both groups was largely Caucasian (Chio, Logroscino et al. 2013), which reflects our patient population. We did not identify an earlier age at disease onset or increased prevalence of bulbar onset as was previously reported for other C9Pos ALS cohorts (Chio,

Borghero et al. 2012, van Rheenen, van Blitterswijk et al. 2012, Irwin, McMillan et al. 2013, Cooper-Knock, Shaw et al. 2014). Bulbar onset was not overrepresented in our C9Pos clinical cohort. C9Pos patients were more likely to report a family history of ALS, and there was an increased prevalence of clinical FTD in this group. We found reduced overall survival in our C9Pos population compared to C9Neg patients is consistent with the published experience from other C9Pos cohorts (Byrne, Elamin et al. 2012, Chio, Borghero et al. 2012, Irwin, McMillan et al. 2013, Borghero, Pugliatti et al. 2014, Montuschi, Iazzolino et al. 2015). We employed the rich resources available through the Emory university brain bank to get a better understanding of pathological differences in C9orf72 patients, compared to sporadic ALS and FTD patients and found that in our small sample p62 immunoreactive regions within the cerebellum are pathognomonic for the C9orf72 expansion genotype. Additionally, paraffin embedded post mortem brain samples were used for immunohistochemical analysis to get a more in depth look at pathological differences across the clinical ALS-FTD disease spectrum. The value of studying these debilitating diseases from human postmortem samples cannot be overstated as animal and cellular models inherently lack the overall complexity of the human brain.

In our analysis of the detergent-insoluble proteome across the ALS-FTD disease spectrum we identified differentially expressed proteins in the insoluble proteome that differentiate ALS/FTD patients from ALS patients. Analyzing the detergent-insoluble proteome provided an enrichment in proteins that aggregate in these diseases and may play important role in disease pathogenesis. Using pairwise correlation coefficients to TDP-43 we were able to make the observation of a strong correlation between TDP-43 and other RNA binding and processing proteins. These

findings provide evidence for the power of proteomics in identifying conserved sets of disease-specific proteins in neurodegenerative conditions.

Finally, we exploit label free mass spectrometry to get a better look at proteins and pathways in the total brain proteome implicated along the ALS-FTD disease spectrum through co-expression network analysis. We found increased inflammation-associated markers in the frontal cortex of ALS/FTD patients compared to patients with only ALS. This observation supports the hypothesis that the dementia phenotype represents changes within the brain that correlate with, if they are not also caused by, increased neuroinflammation (Bettcher and Kramer 2013). ALS patients with the *C9orf72* expansion mutation also had increased inflammation-associated markers in their frontal cortex samples; this cohort is known to develop dementia more frequently than ALS patients without the expansion mutation (Irwin, McMillan et al. 2013, Umoh, Fournier et al. 2016). These findings provide a resource that will invigorate studies moving towards a broad and comprehensive understanding of pathways regulated through biological mechanisms that manifest in co-expression networks with distinguishing characteristics along the ALS-FTD spectrum. Additionally, through orthogonal validation of co-expression modules we have increased confidence in the networks defined using weighted co-expression network analysis in C9Pos ALS frontal cortex brains. This work highlights a multi-omic approach for understanding complex neurodegenerative diseases. The application of a label free proteomic approach to complex neurodegenerative conditions is not original. However, this work represents the first application of this technology across the ALS-FTD disease spectrum and the first application of co-expression analysis to better understand pathways implicated along this disease spectrum.

## **Therapeutic Implications**

The studies performed in this body of work contribute to increasing our understanding of pathways implicated along the ALS-FTD disease spectrum that may serve as potential therapeutic targets. The identification and characterization of overlapping and differentiating pathways affected in this disease spectrum are important in uncovering possible treatment targets. There are currently no treatments available for ALS or FTD. In a review article by Mason et al, they propose a fundamental restructuring of clinical trial design in neurodegenerative diseases - recruitment based on genetic stratification (Mason, Ziemann et al. 2014). They suggest that an essential problem in trial design is that trials mostly recruit patients based on clinical features, such as symptoms and neurological exam findings, and that while this allows maximum study generalizability it assumes most participants share the same underlying disease mechanism or common pathway that the trial drug will converge on. They attest that forming a trial based on presentation of disease is risky because as we and others have shown, patients can share a syndrome that is due to common alterations in brain function but still have dramatically different underlying mechanisms associated with neurodegeneration. An example of this can be seen in our proteomic analysis of C9Pos and C9Neg ALS cases. As we show in our clinical comparison of ALS patients with and without the C9orf72 mutation, there are very few differences in presentation of disease in C9Pos ALS cases compared to sALS cases, however we see a different underlying proteome in these patient brains. Since therapeutic clinical trials are powered based on the assumption that patients will be responsive, drugs targeting specific mechanisms and pathways underlying these clinical phenotypes could keep a drug from affecting a subset of patients and ultimately cause it to be labelled as ineffective (Mason, Ziemann et al. 2014).

## **Future Directions**

While the studies conducted in this work provide evidence of a molecular spectrum of disease and converging and diverging pathways across the ALS-FTD disease spectrum, there is still more work to do to better understand and delineate the pathways involved in these overlapping conditions. In our clinical analysis, we demonstrate few clinical and demographic differences in our C9Pos cohort that exist in the context of significant survival differences. Other studies investigating the relationship between the C9orf72 repeat size and clinical phenotype are needed to show whether the length of the expansion plays a role in disease onset, course, or duration. The limitation in looking at this in our clinical cohort came from technical issues with southern blotting. There are currently only a handful of labs in the country that have successfully sized the repeats using southern blotting techniques. In 2014 in a blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72, Akimoto and colleagues revealed marked differences in results among 14 laboratories (Akimoto, Volk et al. 2014). Due to the wide range of genotyping results they saw in their study, they proposed that southern blotting techniques should be the gold standard and be made obligatory in a clinical diagnostic setting (Akimoto, Volk et al. 2014). However, due to technical limitations this is not commonplace.

Nevertheless, to study whether repeat length acts as a modifying factor in disease, repeat expansion sizing is compulsory. Obtaining the precise size of the expanded repeat is also complicated by several characteristics of the repeat expansion, one of which is somatic

mosaicism that may account for observations of shorter repeat lengths in the blood compared to the frontal cortex (van Blitterswijk, DeJesus-Hernandez et al. 2013, Waite, Baumer et al. 2014, Van Mossevelde, van der Zee et al. 2017). Also, the expansion size in DNA from peripheral blood has been shown to correlate with age at the time of sample collection. Shorter expansion sizes have been noted in the cerebellum relative to DNA from peripheral blood and this could be a consequence of reduced repeat instability in the cerebellum, as some have proposed (Suh, Lee et al. 2015, Van Mossevelde, van der Zee et al. 2017).

There have been very few studies that have tried to explore the relationship between C9orf72 repeat size and disease phenotype. One study found that the length of normal alleles of C9orf72 GGGGCC repeats do not influence disease phenotype (Rutherford, Heckman et al. 2012). While another study showed an inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration (Suh, Lee et al. 2015). A third showed that the C9orf72 repeat size correlates with age of disease onset (Gijssels, Van Mossevelde et al. 2016). The relationship between the expansion length and clinical data could provide critical insights into disease mechanisms and have important implications for diagnostic counseling and therapeutic approaches. Other potential modifiers of C9orf72 expansion should also be explored in the C9Pos ALS clinical cohort that we defined in this work as this represents a substantial north American clinical cohort. One such modifier is Ataxin-2, a gene that codes for a protein involved in endocytosis, modulation of mTOR signaling, modification of ribosomal translation and mitochondrial function (<http://www.genecards.org>) (van Blitterswijk, Mullen et al. 2014). CAG expansion of 35 repeats or more in ataxin2 cause spinocerebellar ataxia 2, while intermediate alleles with 27-33 repeats have been associated with an increased risk for ALS and are a reported

disease modifier in C9orf72 patients (van Blitterswijk, Mullen et al. 2014). Understanding the connection between ATXN2 and the C9orf72 expansion may be another way to increase our understanding of this disease spectrum. These studies are ongoing.

Correlation between TDP-43 and proteins in the detergent-insoluble proteome may point to alterations that may be directly link TDP-43 pathology to neurodegeneration. As such, pathways identified from this analysis could be perturbed in a cellular or animal model of ALS or FTD. Mechanistic studies linking the proteomic changes identified to disease mechanism are needed to strengthen evidence of the causal nature of these correlations.

There are several future directions for the analysis of the total brain proteome. Further studies that tease apart the neuroinflammatory involvement in disease (dementia and c9orf72 related disease) are warranted. Our data point to alterations in several pathways and these can be further teased apart in cellular models. This follow up would not rely on candidate approaches as has been used in the past but instead one could target key drivers of altered co-expression networks and see what effect that has on the network. The benefit of co-expression analysis is that we have presented networks that are linked to clinical phenotypes along the ALS-FTD spectrum which likely reflect underlying molecular alterations. These can be investigated leveraging the networks that we describe. Another future direction for the total proteome analysis is the integration of genetic risks and gene expression networks, as this has been shown to be beneficial in identifying causal mechanisms (Voineagu, Wang et al. 2011, Seyfried, Dammer et al. 2017). Such analysis would strengthen the evidence we have linking certain pathways to this disease spectrum by offering another level of data. This could be done to investigate enrichment of proteins encoded by ALS or FTD risk loci within our total proteome networks. We could possibly extend this to

C9 risk loci as well. The current limitation in this approach lies with the small number of GWAS studies done in ALS/FTD cohorts and their incongruent results. As risk loci for these diseases are more clearly defined these links can be evaluated to see whether any products of GWAS targets are enriched in a particular co-expression module.

Another future direction utilizes the genetic mouse models that have recently been developed to better understand the C9orf72 expansion mutation. Proteomic analysis from disease mice and wildtype mice could be performed and co-expression networks between human and mouse brains can be compared. The biggest challenge with this is the lack of phenotype seen in these C9orf72 mouse models. Though the animals recapitulate disease pathology they rarely display clinical symptoms associated with the ALS-FTD disease spectrum. However, as more models become available that more accurately represent the clinical and pathological features of C9orf72 related disease this could be tackled.

A final future direction of the proteomic work presented in this dissertation relates to the region of the brain studied. For our analyses, we were interested in understanding the overlap of dementia in ALS and the frontal cortex was presumed to be a region that may reflect differences associated with dementia best. Nevertheless, as we are interested in the whole ALS-FTD disease spectrum future studies that look at the motor cortex or spinal cord of patients are reasonable next steps. Co-expression network analysis of proteomic data from these regions would allow us to identify brain region specific networks and we may see similar network alterations or different alterations. ALS is a motor neuron disease, and as such a region with motor neurons should be closely evaluated. Additionally, all of the proteomic analyses conducted in these projects utilized late-stage tissue, future work that incorporate more temporal and dynamic aspects of these

diseases is needed to go from a snapshot to a more comprehensive picture of disease that includes changes over time.

## **Conclusions**

In this study, we coupled label-free mass spectrometry based proteomics and systems biology to define networks of highly correlated proteins associated with neuropathology and cognitive decline in the brains of controls, ALS, ALSFTD, and FTD patients. Pathways altered in ALSFTD patient brains have not previously been explored at a systems biology level, as was carried out in the work of this dissertation. In this work, we address clinical, pathological and proteomic differences in patients along the ALS-FTD disease spectrum with and without the C9orf72 expansion mutation. Future investigations into the proteomic differences identified from this our analyses and their implications in disease are needed to further uncover the specific roles of altered networks in the disease process. These could lead to much needed therapeutic strategies which have been lacking for patients suffering from these conditions. Our findings highlight the utility of large-scale proteomics and integrated systems biology in elucidating the molecular culprits along this disease continuum.

## References

- Abrahams, S., L. H. Goldstein, A. Al-Chalabi, A. Pickering, R. G. Morris, R. E. Passingham, D. J. Brooks and P. N. Leigh (1997). "Relation between cognitive dysfunction and pseudobulbar palsy in amyotrophic lateral sclerosis." J Neurol Neurosurg Psychiatry **62**(5): 464-472.
- Achi, E. Y. and S. A. Rudnicki (2012). "ALS and Frontotemporal Dysfunction: A Review." Neurol Res Int **2012**: 806306.
- Aebersold, R. and M. Mann (2003). "Mass spectrometry-based proteomics." Nature **422**(6928): 198-207.
- Akimoto, C., A. E. Volk, M. van Blitterswijk, M. Van den Broeck, C. S. Leblond, S. Lumbroso, W. Camu, B. Neitzel, O. Onodera, W. van Rheenen, S. Pinto, M. Weber, B. Smith, M. Proven, K. Talbot, P. Keagle, A. Chesi, A. Ratti, J. van der Zee, H. Alstermark, A. Birve, D. Calini, A. Nordin, D. C. Tradowsky, W. Just, H. Daoud, S. Angerbauer, M. DeJesus-Hernandez, T. Konno, A. Lloyd-Jani, M. de Carvalho, K. Mouzat, J. E. Landers, J. H. Veldink, V. Silani, A. D. Gitler, C. E. Shaw, G. A. Rouleau, L. H. van den Berg, C. Van Broeckhoven, R. Rademakers, P. M. Andersen and C. Kubisch (2014). "A blinded international study on the reliability of genetic testing for GGGGCC-repeat expansions in C9orf72 reveals marked differences in results among 14 laboratories." J Med Genet **51**(6): 419-424.
- Al-Chalabi, A., F. Fang, M. F. Hanby, P. N. Leigh, C. E. Shaw, W. Ye and F. Rijdsdijk (2010). "An estimate of amyotrophic lateral sclerosis heritability using twin data." J Neurol Neurosurg Psychiatry **81**(12): 1324-1326.
- Al-Chalabi, A. and O. Hardiman (2013). "The epidemiology of ALS: a conspiracy of genes, environment and time." Nat Rev Neurol **9**(11): 617-628.
- Al-Sarraj, S., A. King, C. Troakes, B. Smith, S. Maekawa, I. Bodi, B. Rogelj, A. Al-Chalabi, T. Hortobagyi and C. E. Shaw (2011). "p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLN and MND/ALS." Acta Neuropathol **122**(6): 691-702.
- Alberici, A., E. Cottini, M. Cosseddu, B. Borroni and A. Padovani (2012). "Suicide risk in frontotemporal lobe degeneration: to be considered, to be prevented." Alzheimer Dis Assoc Disord **26**(2): 194-196.
- Arai, T., M. Hasegawa, H. Akiyama, K. Ikeda, T. Nonaka, H. Mori, D. Mann, K. Tsuchiya, M. Yoshida, Y. Hashizume and T. Oda (2006). "TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis." Biochem Biophys Res Commun **351**(3): 602-611.
- Arnold, E. S., S. C. Ling, S. C. Huelga, C. Lagier-Tourenne, M. Polymenidou, D. Ditsworth, H. B. Kordasiewicz, M. McAlonis-Downes, O. Platoshyn, P. A. Parone, S. Da Cruz, K. M. Clutario, D. Swing, L. Tessarollo, M. Marsala, C. E. Shaw, G. W. Yeo and D. W. Cleveland (2013). "ALS-linked TDP-43 mutations produce aberrant RNA splicing and adult-onset motor neuron disease without aggregation or loss of nuclear TDP-43." Proc Natl Acad Sci U S A **110**(8): E736-745.

Bantscheff, M., M. Schirle, G. Sweetman, J. Rick and B. Kuster (2007). "Quantitative mass spectrometry in proteomics: a critical review." *Anal Bioanal Chem* **389**(4): 1017-1031.

Basso, M., G. Samengo, G. Nardo, T. Massignan, G. D'Alessandro, S. Tartari, L. Cantoni, M. Marino, C. Cheroni, S. De Biasi, M. T. Giordana, M. J. Strong, A. G. Estevez, M. Salmona, C. Bendotti and V. Bonetto (2009). "Characterization of detergent-insoluble proteins in ALS indicates a causal link between oxidative stress and aggregation in pathogenesis." *PLoS One* **4**(12): e8130.

Berrios, G. E. and D. M. Girling (1994). "Introduction: Pick's disease and the 'frontal lobe' dementias." *Hist Psychiatry* **5**(20 Pt 4): 539-547.

Bettcher, B. M. and J. H. Kramer (2013). "Inflammation and clinical presentation in neurodegenerative disease: a volatile relationship." *Neurocase* **19**(2): 182-200.

Bhardwaj, A., M. P. Myers, E. Buratti and F. E. Baralle (2013). "Characterizing TDP-43 interaction with its RNA targets." *Nucleic Acids Res* **41**(9): 5062-5074.

Bi, B., F. Li, J. Guo, C. Li, R. Jing, X. Lv, X. Chen, F. Wang, K. M. Azadzi, L. Wang, Y. Liu and J. H. Yang (2017). "Label-free quantitative proteomics unravels the importance of RNA processing in glioma malignancy." *Neuroscience*.

Bjørkøy, G., T. Lamark, S. Pankiv, A. Øvervatn, A. Brech and T. Johansen (2009). "Chapter 12: Monitoring Autophagic Degradation of p62/SQSTM1." *Methods in Enzymology* **452**: 181-197.

Blair, I. P., K. L. Williams, S. T. Warraich, J. C. Durnall, A. D. Thoeng, J. Manavis, P. C. Blumbergs, S. Vucic, M. C. Kiernan and G. A. Nicholson (2010). "FUS mutations in amyotrophic lateral sclerosis: clinical, pathological, neurophysiological and genetic analysis." *J Neurol Neurosurg Psychiatry* **81**(6): 639-645.

Blokhuis, A. M., M. Koppers, E. J. Groen, D. M. van den Heuvel, S. Dini Modigliani, J. J. Anink, K. Fumoto, F. van Diggelen, A. Sneltink, P. Sogaar, B. M. Verheijen, J. A. Demmers, J. H. Veldink, E. Aronica, I. Bozzoni, J. den Hertog, L. H. van den Berg and R. J. Pasterkamp (2016). "Comparative interactomics analysis of different ALS-associated proteins identifies converging molecular pathways." *Acta Neuropathol* **132**(2): 175-196.

Boeve, B. F., K. B. Boylan, N. R. Graff-Radford, M. DeJesus-Hernandez, D. S. Knopman, O. Pedraza, P. Vemuri, D. Jones, V. Lowe, M. E. Murray, D. W. Dickson, K. A. Josephs, B. K. Rush, M. M. Machulda, J. A. Fields, T. J. Ferman, M. Baker, N. J. Rutherford, J. Adamson, Z. K. Wszolek, A. Adeli, R. Savica, B. Boot, K. M. Kuntz, R. Gavrilo, A. Reeves, J. Whitwell, K. Kantarci, C. R. Jack, J. E. Parisi, J. A. Lucas, R. C. Petersen and R. Rademakers (2012). "Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72." *Brain* **135**(3): 765-783.

Borghero, G., M. Pugliatti, F. Marrosu, M. G. Marrosu, M. R. Murru, G. Floris, A. Cannas, L. D. Parish, P. Occhineri, T. B. Cau, D. Loi, A. Ticca, S. Traccis, U. Manera, A. Canosa, C. Moglia, A. Calvo, M. Barberis, M. Brunetti, H. A. Pliner, A. E. Renton, M. A. Nalls, B. J. Traynor, G. Restagno and A. Chio (2014). "Genetic architecture of ALS in Sardinia." *Neurobiol Aging* **35**(12): 2882.e2887-2882.e2812.

Braak, H., A. Ludolph, D. R. Thal and K. Del Tredici (2010). "Amyotrophic lateral sclerosis: dash-like accumulation of phosphorylated TDP-43 in somatodendritic and axonal compartments

of somatomotor neurons of the lower brainstem and spinal cord." Acta Neuropathol **120**(1): 67-74.

Brettschneider, J., K. Del Tredici, D. J. Irwin, M. Grossman, J. L. Robinson, J. B. Toledo, E. B. Lee, L. Fang, V. M. Van Deerlin, A. C. Ludolph, V. M. Lee, H. Braak and J. Q. Trojanowski (2014). "Sequential distribution of pTDP-43 pathology in behavioral variant frontotemporal dementia (bvFTD)." Acta Neuropathol **127**(3): 423-439.

Brettschneider, J., J. B. Toledo, V. M. Van Deerlin, L. Elman, L. McCluskey, V. M. Lee and J. Q. Trojanowski (2012). "Microglial activation correlates with disease progression and upper motor neuron clinical symptoms in amyotrophic lateral sclerosis." PLoS One **7**(6): e39216.

Brooks, B. R. (1994). "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 **124 Suppl**: 96-107.

Brooks, B. R., R. G. Miller, M. Swash and T. L. Munsat (2000). "El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis." Amyotroph Lateral Scler Other Motor Neuron Disord **1**(5): 293-299.

Buratti, E. and F. E. Baralle (2010). "The multiple roles of TDP-43 in pre-mRNA processing and gene expression regulation." RNA Biol **7**(4): 420-429.

Burberry, A., N. Suzuki, J. Y. Wang, R. Moccia, D. A. Mordes, M. H. Stewart, S. Suzuki-Uematsu, S. Ghosh, A. Singh, F. T. Merkle, K. Koszka, Q. Z. Li, L. Zon, D. J. Rossi, J. J. Trowbridge, L. D. Notarangelo and K. Eggan (2016). "Loss-of-function mutations in the C9ORF72 mouse ortholog cause fatal autoimmune disease." Sci Transl Med **8**(347): 347ra393.

Burns, T. C., M. D. Li, S. Mehta, A. J. Awad and A. A. Morgan (2015). "Mouse models rarely mimic the transcriptome of human neurodegenerative diseases: A systematic bioinformatics-based critique of preclinical models." Eur J Pharmacol **759**: 101-117.

Byrne, S., M. Elamin, P. Bede, A. Shatunov, C. Walsh, B. Corr, M. Heverin, N. Jordan, K. Kenna, C. Lynch, R. L. McLaughlin, P. M. Iyer, C. O'Brien, J. Phukan, B. Wynne, A. L. Bokde, D. G. Bradley, N. Pender, A. Al-Chalabi and O. Hardiman (2012). "Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a C9orf72 repeat expansion: a population-based cohort study." Lancet Neurol **11**(3): 232-240.

Callister, S. J., R. C. Barry, J. N. Adkins, E. T. Johnson, W. J. Qian, B. J. Webb-Robertson, R. D. Smith and M. S. Lipton (2006). "Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics." J Proteome Res **5**(2): 277-286.

Carvey, P. M., B. Hendey and A. J. Monahan (2009). "The Blood Brain Barrier in Neurodegenerative Disease: A Rhetorical Perspective." Journal of neurochemistry **111**(2): 291-314.

Chen, Y., X. H. Liu, J. J. Wu, H. M. Ren, J. Wang, Z. T. Ding and Y. P. Jiang (2016). "Proteomic analysis of cerebrospinal fluid in amyotrophic lateral sclerosis." Exp Ther Med **11**(6): 2095-2106.

Chio, A., G. Borghero, G. Restagno, G. Mora, C. Drepper, B. J. Traynor, M. Sendtner, M. Brunetti, I. Ossola, A. Calvo, M. Pugliatti, M. A. Sotgiu, M. R. Murru, M. G. Marrosu, F. Marrosu, K. Marinou, J. Mandrioli, P. Sola, C. Caponnetto, G. Mancardi, P. Mandich, V. La Bella, R. Spataro, A. Conte, M. R. Monsurro, G. Tedeschi, F. Pisano, I. Bartolomei, F. Salvi, G. Lauria Pinter, I. Simone, G. Logroscino, A. Gambardella, A. Quattrone, C. Lunetta, P. Volanti, M. Zollino, S. Penco, S. Battistini, I. consortium, A. E. Renton, E. Majounie, Y. Abramzon, F. L. Conforti, F. Giannini, M. Corbo and M. Sabatelli (2012). "Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72." *Brain* **135**(Pt 3): 784-793.

Chio, A., G. Logroscino, O. Hardiman, R. Swingler, D. Mitchell, E. Beghi and B. G. Traynor (2009). "Prognostic factors in ALS: A critical review." *Amyotroph Lateral Scler* **10**(5-6): 310-323.

Chio, A., G. Logroscino, B. J. Traynor, J. Collins, J. C. Simeone, L. A. Goldstein and L. A. White (2013). "Global epidemiology of amyotrophic lateral sclerosis: a systematic review of the published literature." *Neuroepidemiology* **41**(2): 118-130.

Collins, M. A., J. An, B. L. Hood, T. P. Conrads and R. P. Bowser (2015). "Label-Free LC-MS/MS Proteomic Analysis of Cerebrospinal Fluid Identifies Protein/Pathway Alterations and Candidate Biomarkers for Amyotrophic Lateral Sclerosis." *J Proteome Res* **14**(11): 4486-4501.

Conforti, F. L., R. Barone, S. L. Fermo, C. Giliberto, F. Patti, A. Gambardella, A. Quattrone and M. Zappia (2011). "Sporadic motor neuron disease in a familial novel SOD1 mutation: incomplete penetrance or chance association?" *Amyotroph Lateral Scler* **12**(3): 220-222.

Conlon, E. G., L. Lu, A. Sharma, T. Yamazaki, T. Tang, N. A. Shneider and J. L. Manley (2016). "The C9ORF72 GGGGCC expansion forms RNA G-quadruplex inclusions and sequesters hnRNP H to disrupt splicing in ALS brains." *Elife* **5**.

Cooper-Knock, J., C. Hewitt, J. R. Highley, A. Brockington, A. Milano, S. Man, J. Martindale, J. Hartley, T. Walsh, C. Gelsthorpe, L. Baxter, G. Forster, M. Fox, J. Bury, K. Mok, C. J. McDermott, B. J. Traynor, J. Kirby, S. B. Wharton, P. G. Ince, J. Hardy and P. J. Shaw (2012). "Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72." *Brain* **135**(Pt 3): 751-764.

Cooper-Knock, J., P. J. Shaw and J. Kirby (2014). "The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype." *Acta Neuropathol* **127**(3): 333-345.

Cooper-Knock, J., M. J. Walsh, A. Higginbottom, J. Robin Highley, M. J. Dickman, D. Edbauer, P. G. Ince, S. B. Wharton, S. A. Wilson, J. Kirby, G. M. Hautbergue and P. J. Shaw (2014). "Sequestration of multiple RNA recognition motif-containing proteins by C9orf72 repeat expansions." *Brain* **137**(Pt 7): 2040-2051.

Cox, J., M. Y. Hein, C. A. Luber, I. Paron, N. Nagaraj and M. Mann (2014). "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." *Mol Cell Proteomics* **13**(9): 2513-2526.

Dammer, E. B., D. M. Duong, I. Diner, M. Gearing, Y. Feng, J. J. Lah, A. I. Levey and N. T. Seyfried (2013). "Neuron enriched nuclear proteome isolated from human brain." *J Proteome Res* **12**(7): 3193-3206.

Dammer, E. B., C. Fallini, Y. M. Gozal, D. M. Duong, W. Rossoll, P. Xu, J. J. Lah, A. I. Levey, J. Peng, G. J. Bassell and N. T. Seyfried (2012). "Coaggregation of RNA-binding proteins in a model of TDP-43 proteinopathy with selective RGG motif methylation and a role for RRM1 ubiquitination." PLoS One **7**(6): e38658.

Dammer, E. B., A. K. Lee, D. M. Duong, M. Gearing, J. J. Lah, A. I. Levey and N. T. Seyfried (2015). "Quantitative Phosphoproteomics of Alzheimer's Disease Reveals Crosstalk between Kinases and Small Heat Shock Proteins." Proteomics **15**(0): 508-519.

De Strooper, B. and E. Karran (2016). "The Cellular Phase of Alzheimer's Disease." Cell **164**(4): 603-615.

DeJesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, A. M. Nicholson, N. A. Finch, H. Flynn, J. Adamson, N. Kouri, A. Wojtas, P. Sengdy, G. Y. Hsiung, A. Karydas, W. W. Seeley, K. A. Josephs, G. Coppola, D. H. Geschwind, Z. K. Wszolek, H. Feldman, D. S. Knopman, R. C. Petersen, B. L. Miller, D. W. Dickson, K. B. Boylan, N. R. Graff-Radford and R. Rademakers (2011). "Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS." Neuron **72**(2): 245-256.

del Aguila, M. A., W. T. Longstreth, Jr., V. McGuire, T. D. Koepsell and G. van Belle (2003). "Prognosis in amyotrophic lateral sclerosis: a population-based study." Neurology **60**(5): 813-819.

Deng, H. X., W. Chen, S. T. Hong, K. M. Boycott, G. H. Gorrie, N. Siddique, Y. Yang, F. Fecto, Y. Shi, H. Zhai, H. Jiang, M. Hirano, E. Rampersaud, G. H. Jansen, S. Donkervoort, E. H. Bigio, B. R. Brooks, K. Ajroud, R. L. Sufit, J. L. Haines, E. Mugnaini, M. A. Pericak-Vance and T. Siddique (2011). "Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia." Nature **477**(7363): 211-215.

Devenney, E., S. Vucic, J. R. Hodges and M. C. Kiernan (2015). "Motor neuron disease-frontotemporal dementia: a clinical continuum." Expert Rev Neurother **15**(5): 509-522.

Dickson, D. W. (1998). "Pick's disease: a modern approach." Brain Pathol **8**(2): 339-354.

Donnelly, C. J., J. C. Grima and R. Sattler (2014). "Aberrant RNA homeostasis in amyotrophic lateral sclerosis: potential for new therapeutic targets?" Neurodegener Dis Manag **4**(6): 417-437.

Donovan, L., E. Dammer, D. Duong, J. Hanfelt, A. Levey, N. Seyfried and J. Lah (2013). "Exploring the potential of the platelet membrane proteome as a source of peripheral biomarkers for Alzheimer's disease." Alzheimer's Research & Therapy **5**(3): 32.

Donovan, L. E., L. Higginbotham, E. B. Dammer, M. Gearing, H. D. Rees, Q. Xia, D. M. Duong, N. T. Seyfried, J. J. Lah and A. I. Levey (2012). "Analysis of a membrane-enriched proteome from postmortem human brain tissue in Alzheimer's disease." Proteomics Clin Appl **6**(3-4): 201-211.

Dugger, B. N. and D. W. Dickson (2017). "Pathology of Neurodegenerative Diseases." Cold Spring Harb Perspect Biol.

Elias, J. E. and S. P. Gygi (2007). "Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry." Nat Meth **4**(3): 207-214.

- Engelen-Lee, J., A. M. Blokhuis, W. G. Spliet, R. J. Pasterkamp, E. Aronica, J. A. Demmers, R. Broekhuizen, G. Nardo, N. Bovenschen and L. H. Van Den Berg (2016). "Proteomic profiling of the spinal cord in ALS: decreased ATP5D levels suggest synaptic dysfunction in ALS pathogenesis." *Amyotroph Lateral Scler Frontotemporal Degener*: 1-11.
- Engelhart, M., M. Geerlings, J. Meijer, A. Kiliaan, A. Ruitenber, J. van Swieten, T. Stijnen, A. Hofman, J. Witteman and M. Breteler (2004). "Inflammatory proteins in plasma and the risk of dementia: The rotterdam study." *Arch Neurol* **61**.
- Farg, M. A., V. Sundaramoorthy, J. M. Sultana, S. Yang, R. A. Atkinson, V. Levina, M. A. Halloran, P. A. Gleeson, I. P. Blair, K. Y. Soo, A. E. King and J. D. Atkin (2014). "C9ORF72, implicated in amyotrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking." *Hum Mol Genet* **23**(13): 3579-3595.
- Farrelly, L. A., M. Föcking and D. R. Cotter (2014). Insights from Proteomic Studies on Schizophrenia Preclinical Models: What Can We Learn for Drug Discovery? *Proteomics and Metabolomics in Psychiatry*. M.-d.-S. D. Department of Psychiatry, Royal College of Surgeons in Ireland, Smurfit Education and Research Centre, Beaumont Hospital, Dublin, Ireland, Basel, Karger. **29**: 56-73.
- Fecto, F. and T. Siddique (2011). "Making connections: pathology and genetics link amyotrophic lateral sclerosis with frontotemporal lobe dementia." *J Mol Neurosci* **45**(3): 663-675.
- Ferrari, R., D. Kapogiannis, E. D. Huey and P. Momeni (2011). "FTD and ALS: a tale of two diseases." *Curr Alzheimer Res* **8**(3): 273-294.
- Finlayson, M. H. and J. B. Martin (1973). "Cerebral lesions in familial amyotrophic lateral sclerosis and dementia." *Acta Neuropathol* **26**(3): 237-246.
- Floeter, M. K., D. Bageac, L. E. Danielian, L. E. Braun, B. J. Traynor and J. Y. Kwan (2016). "Longitudinal imaging in C9orf72 mutation carriers: Relationship to phenotype." *Neuroimage Clin* **12**: 1035-1043.
- Fong, J. C., A. M. Karydas and J. S. Goldman (2012). "Genetic counseling for FTD/ALS caused by the C9ORF72 hexanucleotide expansion." *Alzheimers Res Ther* **4**(4): 27.
- Frackman, S., G. Kobs, D. Simpson and D. Storts (1998). Betaine and DMSO: Enhancing Agents for PCR. *Promega Notes* Promega. **65**: 27.
- Freer, C., T. Hylton, H. M. Jordan, W. E. Kaye, S. Singh and Y. Huang (2015). "Results of Florida's Amyotrophic Lateral Sclerosis Surveillance Project, 2009-2011." *BMJ Open* **5**(4): e007359.
- Freibaum, B. D., R. Chitta, A. A. High and J. P. Taylor (2010). "Global analysis of TDP-43 interacting proteins reveals strong association with RNA splicing and translation machinery." *Journal of proteome research* **9**(2): 1104-1120.
- Freibaum, B. D., Y. Lu, R. Lopez-Gonzalez, N. C. Kim, S. Almeida, K. H. Lee, N. Badders, M. Valentine, B. L. Miller, P. C. Wong, L. Petrucelli, H. J. Kim, F. B. Gao and J. P. Taylor (2015). "GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport." *Nature* **525**(7567): 129-133.

Gaiteri, C., Y. Ding, B. French, G. C. Tseng and E. Sibille (2014). "Beyond modules and hubs: the potential of gene coexpression networks for investigating molecular mechanisms of complex brain disorders." Genes Brain Behav **13**(1): 13-24.

Galimberti, D., B. Arosio, C. Fenoglio, M. Serpente, S. M. Cioffi, R. Bonsi, P. Rossi, C. Abbate, D. Mari and E. Scarpini (2014). "Incomplete penetrance of the C9ORF72 hexanucleotide repeat expansions: frequency in a cohort of geriatric non-demented subjects." J Alzheimers Dis **39**(1): 19-22.

Garcia-Redondo, A., O. Dols-Icardo, R. Rojas-Garcia, J. Esteban-Perez, P. Cordero-Vazquez, J. L. Munoz-Blanco, I. Catalina, M. Gonzalez-Munoz, L. Varona, E. Sarasola, M. Povedano, T. Sevilla, A. Guerrero, J. Pardo, A. Lopez de Munain, C. Marquez-Infante, F. J. de Rivera, P. Pastor, I. Jerico, A. A. de Arcaya, J. S. Mora, J. Clarimon, J. F. Gonzalo-Martinez, A. Juarez-Rufian, G. Atencia, R. Jimenez-Bautista, Y. Moran, J. Mascias, M. Hernandez-Barral, S. Kapetanovic, M. Garcia-Barcina, C. Alcalá, A. Vela, C. Ramirez-Ramos, L. Galan, J. Perez-Tur, B. Quintans, M. J. Sobrido, R. Fernandez-Torron, J. J. Poza, A. Gorostidi, C. Paradas, P. Villoslada, P. Larrode, J. L. Capablo, J. Pascual-Calvet, M. Goni, Y. Morgado, M. Guitart, S. Moreno-Laguna, A. Rueda, C. Martin-Estefania, C. Cemillan, R. Blesa and A. Lleo (2013). "Analysis of the C9orf72 gene in patients with amyotrophic lateral sclerosis in Spain and different populations worldwide." Hum Mutat **34**(1): 79-82.

Gendron, T. F., J. Chew, J. N. Stankowski, L. R. Hayes, Y. J. Zhang, M. Prudencio, Y. Carlomagno, L. M. Daugherty, K. Jansen-West, E. A. Perkerson, A. O'Raw, C. Cook, L. Pregent, V. Belzil, M. van Blitterswijk, L. J. Tabassian, C. W. Lee, M. Yue, J. Tong, Y. Song, M. Castanedes-Casey, L. Rousseau, V. Phillips, D. W. Dickson, R. Rademakers, J. D. Fryer, B. K. Rush, O. Pedraza, A. M. Caputo, P. Desaro, C. Palmucci, A. Robertson, M. G. Heckman, N. N. Diehl, E. Wiggs, M. Tierney, L. Braun, J. Farren, D. Lacomis, S. Ladha, C. N. Fournier, L. F. McCluskey, L. B. Elman, J. B. Toledo, J. D. McBride, C. Tiloca, C. Morelli, B. Poletti, F. Solca, A. Prella, J. Wu, J. Jockel-Balsarotti, F. Rigo, C. Ambrose, A. Datta, W. Yang, D. Raitcheva, G. Antognetti, A. McCampbell, J. C. Van Swieten, B. L. Miller, A. L. Boxer, R. H. Brown, R. Bowser, T. M. Miller, J. Q. Trojanowski, M. Grossman, J. D. Berry, W. T. Hu, A. Ratti, B. J. Traynor, M. D. Disney, M. Benatar, V. Silani, J. D. Glass, M. K. Floeter, J. D. Rothstein, K. B. Boylan and L. Petrucelli (2017). "Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis." Sci Transl Med **9**(383).

Geser, F., V. M. Lee and J. Q. Trojanowski (2010). "Amyotrophic lateral sclerosis and frontotemporal lobar degeneration: a spectrum of TDP-43 proteinopathies." Neuropathology **30**(2): 103-112.

Gibbs, D. L., A. Baratt, R. S. Baric, Y. Kawaoka, R. D. Smith, E. S. Orwoll, M. G. Katze and S. K. McWeeny (2013). "Protein co-expression network analysis (ProCoNA)." J Clin Bioinforma **3**(1): 11.

Gijssels, I., S. Van Mossevelde, J. van der Zee, A. Sieben, S. Engelborghs, J. De Bleeker, A. Ivanoiu, O. Deryck, D. Edbauer, M. Zhang, B. Heeman, V. Baumer, M. Van den Broeck, M. Mattheijssens, K. Peeters, E. Rogaeva, P. De Jonghe, P. Cras, J. J. Martin, P. P. de Deyn, M. Cruts and C. Van Broeckhoven (2016). "The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter." Mol Psychiatry **21**(8): 1112-1124.

- Gitler, A. D. and H. Tsuiji (2016). "There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS." Brain Res **1647**: 19-29.
- Goedert, M., R. A. Crowther and M. G. Spillantini (1998). "Tau mutations cause frontotemporal dementias." Neuron **21**(5): 955-958.
- Goetz, C. G. (2000). "AMYOTROPHIC LATERAL SCLEROSIS: EARLY CONTRIBUTIONS OF JEAN-MARTIN CHARCOT." Muscle & Nerve **23**: 336-343.
- Gozal, Y. M., D. M. Duong, M. Gearing, D. Cheng, J. J. Hanfelt, C. Funderburk, J. Peng, J. J. Lah and A. I. Levey (2009). "Proteomics analysis reveals novel components in the detergent-insoluble subproteome in Alzheimer's disease." J Proteome Res **8**(11): 5069-5079.
- Gozal, Y. M., D. M. Duong, M. Gearing, D. Cheng, J. J. Hanfelt, C. Funderburk, J. Peng, J. J. Lah and A. I. Levey (2009). "Proteomics analysis reveals novel components in the detergent-insoluble subproteome in Alzheimer's disease." Journal of Proteome Research **8**(11): 5069-5079.
- Greenway, M. J., P. M. Andersen, C. Russ, S. Ennis, S. Cashman, C. Donaghy, V. Patterson, R. Swingler, D. Kieran, J. Prehn, K. E. Morrison, A. Green, K. R. Acharya, R. H. Brown, Jr. and O. Hardiman (2006). "ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis." Nat Genet **38**(4): 411-413.
- Gstaiger, M. and R. Aebersold (2009). "Applying mass spectrometry-based proteomics to genetics, genomics and network biology." Nat Rev Genet **10**(9): 617-627.
- Guido, N., E. Starostina, D. Leake and I. Saaem (2016). "Improved PCR Amplification of Broad Spectrum GC DNA Templates." PLoS One **11**(6): e0156478.
- Hales, C. M., E. B. Dammer, Q. Deng, D. M. Duong, M. Gearing, J. C. Troncoso, M. Thambisetty, J. J. Lah, J. M. Shulman, A. I. Levey and N. T. Seyfried (2016). "Changes in the detergent-insoluble brain proteome linked to amyloid and tau in Alzheimer's Disease progression." Proteomics **16**(23): 3042-3053.
- Hodges, J. (2012). "Familial frontotemporal dementia and amyotrophic lateral sclerosis associated with the C9ORF72 hexanucleotide repeat." Brain **135**(Pt 3): 652-655.
- Hodges, J. R., R. Davies, J. Xuereb, J. Kril and G. Halliday (2003). "Survival in frontotemporal dementia." Neurology **61**(3): 349-354.
- Hovhannisyan, R. H. and R. P. Carstens (2007). "Heterogeneous ribonucleoprotein m is a splicing regulatory protein that can enhance or silence splicing of alternatively spliced exons." J Biol Chem **282**(50): 36265-36274.
- Hsiung, G. Y., M. DeJesus-Hernandez, H. H. Feldman, P. Sengdy, P. Bouchard-Kerr, E. Dwosh, R. Butler, B. Leung, A. Fok, N. J. Rutherford, M. Baker, R. Rademakers and I. R. Mackenzie (2012). "Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p." Brain **135**(Pt 3): 709-722.
- Hu, W. T., M. Shelnett, A. Wilson, N. Yarab, C. Kelly, M. Grossman, D. J. Libon, J. Khan, J. J. Lah, A. I. Levey and J. Glass (2013). "Behavior matters--cognitive predictors of survival in amyotrophic lateral sclerosis." PLoS One **8**(2): e57584.
- Hudson, A. J. (1981). "Amyotrophic lateral sclerosis and its association with dementia, parkinsonism and other neurological disorders: a review." Brain **104**(2): 217-247.

- Ideker, T., V. Thorsson, J. A. Ranish, R. Christmas, J. Buhler, J. K. Eng, R. Bumgarner, D. R. Goodlett, R. Aebersold and L. Hood (2001). "Integrated genomic and proteomic analyses of a systematically perturbed metabolic network." Science **292**(5518): 929-934.
- Igaz, L. M., L. K. Kwong, Y. Xu, A. C. Truax, K. Uryu, M. Neumann, C. M. Clark, L. B. Elman, B. L. Miller, M. Grossman, L. F. McCluskey, J. Q. Trojanowski and V. M. Lee (2008). "Enrichment of C-terminal fragments in TAR DNA-binding protein-43 cytoplasmic inclusions in brain but not in spinal cord of frontotemporal lobar degeneration and amyotrophic lateral sclerosis." Am J Pathol **173**(1): 182-194.
- Ingre, C., P. M. Roos, F. Piehl, F. Kamel and F. Fang (2015). "Risk factors for amyotrophic lateral sclerosis." Clin Epidemiol **7**: 181-193.
- Irwin, D. J., C. T. McMillan, J. Brettschneider, D. J. Libon, J. Powers, K. Rascovsky, J. B. Toledo, A. Boller, J. Bekisz, K. Chandrasekaran, E. M. Wood, L. M. Shaw, J. H. Woo, P. A. Cook, D. A. Wolk, S. E. Arnold, V. M. Van Deerlin, L. F. McCluskey, L. Elman, V. M. Lee, J. Q. Trojanowski and M. Grossman (2013). "Cognitive decline and reduced survival in C9orf72 expansion frontotemporal degeneration and amyotrophic lateral sclerosis." J Neurol Neurosurg Psychiatry **84**(2): 163-169.
- Ivetic, A. and A. J. Ridley (2004). "Ezrin/radixin/moesin proteins and Rho GTPase signalling in leucocytes." Immunology **112**(2): 165-176.
- James, P. A. and K. Talbot (2006). "The molecular genetics of non-ALS motor neuron diseases." Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease **1762**(11-12): 986-1000.
- Janssens, J. and C. Van Broeckhoven (2013). "Pathological mechanisms underlying TDP-43 driven neurodegeneration in FTL-ALS spectrum disorders." Hum Mol Genet.
- Jones, C. T., D. J. Brock, A. M. Chancellor, C. P. Warlow and R. J. Swingle (1993). "Cu/Zn superoxide dismutase (SOD1) mutations and sporadic amyotrophic lateral sclerosis." Lancet **342**(8878): 1050-1051.
- Kabashi, E., P. N. Valdmanis, P. Dion, D. Spiegelman, B. J. McConkey, C. Vande Velde, J. P. Bouchard, L. Lacomblez, K. Pochigava, F. Salachas, P. F. Pradat, W. Camu, V. Meininger, N. Dupre and G. A. Rouleau (2008). "TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis." Nat Genet **40**(5): 572-574.
- Kafasla, P., M. Patrino-Georgoula, J. D. Lewis and A. Guialis (2002). "Association of the 72/74-kDa proteins, members of the heterogeneous nuclear ribonucleoprotein M group, with the pre-mRNA at early stages of spliceosome assembly." Biochem J **363**(Pt 3): 793-799.
- Karpievitch, Y. V., A. R. Dabney and R. D. Smith (2012). "Normalization and missing value imputation for label-free LC-MS analysis." BMC Bioinformatics **13**(Suppl 16): S5-S5.
- Karpievitch, Y. V., A. R. Dabney and R. D. Smith (2012). "Normalization and missing value imputation for label-free LC-MS analysis." BMC Bioinformatics **13** Suppl 16: S5.
- Kersaitis, C., G. M. Halliday and J. J. Kril (2004). "Regional and cellular pathology in frontotemporal dementia: relationship to stage of disease in cases with and without Pick bodies." Acta Neuropathol **108**(6): 515-523.

- Kiernan, M. C., S. Vucic, B. C. Cheah, M. R. Turner, A. Eisen, O. Hardiman, J. R. Burrell and M. C. Zoing (2011). "Amyotrophic lateral sclerosis." Lancet **377**(9769): 942-955.
- Kwon, I., S. Xiang, M. Kato, L. Wu, P. Theodoropoulos, T. Wang, J. Kim, J. Yun, Y. Xie and S. L. McKnight (2014). "Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells." Science **345**(6201): 1139-1145.
- Landqvist Waldo, M., A. F. Santillo, L. Gustafson, E. Englund and U. Passant (2014). "Somatic complaints in frontotemporal dementia." Am J Neurodegener Dis **3**(2): 84-92.
- Langfelder, P. and S. Horvath (2012). "Fast R Functions for Robust Correlations and Hierarchical Clustering." Journal of statistical software **46**(11): i11.
- Langfelder, P. and S. Horvath (2012). "Fast R Functions for Robust Correlations and Hierarchical Clustering." J Stat Softw **46**(11).
- Langfelder, P., B. Zhang and S. Horvath (2008). "Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R." Bioinformatics **24**(5): 719-720.
- Lee, E. B., V. M. Lee and J. Q. Trojanowski (2012). "Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration." Nat Rev Neurosci **13**(1): 38-50.
- Lee, Y. B., H. J. Chen, J. N. Peres, J. Gomez-Deza, J. Attig, M. Stalekar, C. Troakes, A. L. Nishimura, E. L. Scotter, C. Vance, Y. Adachi, V. Sardone, J. W. Miller, B. N. Smith, J. M. Gallo, J. Ule, F. Hirth, B. Rogelj, C. Houart and C. E. Shaw (2013). "Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic." Cell Rep **5**(5): 1178-1186.
- Leigh, P. N., B. H. Anderton, A. Dodson, J. M. Gallo, M. Swash and D. M. Power (1988). "Ubiquitin deposits in anterior horn cells in motor neurone disease." Neurosci Lett **93**(2-3): 197-203.
- Leigh, P. N., H. Whitwell, O. Garofalo, J. Buller, M. Swash, J. E. Martin, J. M. Gallo, R. O. Weller and B. H. Anderton (1991). "Ubiquitin-immunoreactive intraneuronal inclusions in amyotrophic lateral sclerosis. Morphology, distribution, and specificity." Brain **114** ( Pt 2): 775-788.
- Li, Y., M. Collins, J. An, R. Geiser, T. Tegeler, K. Tsantilas, K. Garcia, P. Pirrotte and R. Bowser (2016). "Immunoprecipitation and mass spectrometry defines an extensive RBM45 protein-protein interaction network." Brain Res **1647**: 79-93.
- Lomen-Hoerth, C. (2011). "Clinical phenomenology and neuroimaging correlates in ALS-FTD." J Mol Neurosci **45**(3): 656-662.
- Lomen-Hoerth, C., T. Anderson and B. Miller (2002). "The overlap of amyotrophic lateral sclerosis and frontotemporal dementia." Neurology **59**(7): 1077-1079.
- Lomen-Hoerth, C., J. Murphy, S. Langmore, J. H. Kramer, R. K. Olney and B. Miller (2003). "Are amyotrophic lateral sclerosis patients cognitively normal?" Neurology **60**(7): 1094-1097.
- Lopez-Gonzalez, R., Y. Lu, T. F. Gendron, A. Karydas, H. Tran, D. Yang, L. Petrucelli, B. L. Miller, S. Almeida and F. B. Gao (2016). "Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons." Neuron **92**(2): 383-391.

- Luber, C. A., J. Cox, H. Lauterbach, B. Fancke, M. Selbach, J. Tschopp, S. Akira, M. Wiegand, H. Hochrein, M. O'Keeffe and M. Mann (2010). "Quantitative proteomics reveals subset-specific viral recognition in dendritic cells." *Immunity* **32**(2): 279-289.
- Lundgren, D. H., S. I. Hwang, L. Wu and D. K. Han (2010). "Role of spectral counting in quantitative proteomics." *Expert Rev Proteomics* **7**(1): 39-53.
- Mackenzie, I. R., M. Baker, S. Pickering-Brown, G. Y. Hsiung, C. Lindholm, E. Dwosh, J. Gass, A. Cannon, R. Rademakers, M. Hutton and H. H. Feldman (2006). "The neuropathology of frontotemporal lobar degeneration caused by mutations in the progranulin gene." *Brain* **129**(Pt 11): 3081-3090.
- Mackenzie, I. R. and R. Rademakers (2008). "The role of transactive response DNA-binding protein-43 in amyotrophic lateral sclerosis and frontotemporal dementia." *Curr Opin Neurol* **21**(6): 693-700.
- Magnus, T., M. Beck, R. Giess, I. Puls, M. Naumann and K. V. Toyka (2002). "Disease progression in amyotrophic lateral sclerosis: predictors of survival." *Muscle Nerve* **25**(5): 709-714.
- Mahoney, C. J., J. Beck, J. D. Rohrer, T. Lashley, K. Mok, T. Shakespeare, T. Yeatman, E. K. Warrington, J. M. Schott, N. C. Fox, M. N. Rossor, J. Hardy, J. Collinge, T. Revesz, S. Mead and J. D. Warren (2012). "Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features." *Brain* **135**(Pt 3): 736-750.
- Majounie, E., A. E. Renton, K. Mok, E. G. P. Dopper, A. Waite, S. Rollinson, A. Chiò, G. Restagno, N. Nicolaou, J. Simon-Sanchez, J. C. van Swieten, Y. Abramzon, J. O. Johnson, M. Sendtner, R. Pamphlett, R. W. Orrell, S. Mead, K. C. Sidle, H. Houlden, J. D. Rohrer, K. E. Morrison, H. Pall, K. Talbot, O. Ansorge, D. G. Hernandez, S. Arepalli, M. Sabatelli, G. Mora, M. Corbo, F. Giannini, A. Calvo, E. Englund, G. Borghero, G. L. Floris, A. M. Remes, H. Laaksovirta, L. McCluskey, J. Q. Trojanowski, V. M. Van Deerlin, G. D. Schellenberg, M. A. Nalls, V. E. Drory, C.-S. Lu, T.-H. Yeh, H. Ishiura, Y. Takahashi, S. Tsuji, I. Le Ber, A. Brice, C. Drepper, N. Williams, J. Kirby, P. Shaw, J. Hardy, P. J. Tienari, P. Heutink, H. R. Morris, S. Pickering-Brown and B. J. Traynor (2012). "Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study." *The Lancet Neurology* **11**(4): 323-330.
- Marko, M., M. Leichter, M. Patrino-Georgoula and A. Guialis (2014). "Selective interactions of hnRNP M isoforms with the TET proteins TAF15 and TLS/FUS." *Mol Biol Rep* **41**(4): 2687-2695.
- Maruyama, H., H. Morino, H. Ito, Y. Izumi, H. Kato, Y. Watanabe, Y. Kinoshita, M. Kamada, H. Nodera, H. Suzuki, O. Komure, S. Matsuura, K. Kobatake, N. Morimoto, K. Abe, N. Suzuki, M. Aoki, A. Kawata, T. Hirai, T. Kato, K. Ogasawara, A. Hirano, T. Takumi, H. Kusaka, K. Hagiwara, R. Kaji and H. Kawakami (2010). "Mutations of optineurin in amyotrophic lateral sclerosis." *Nature* **465**(7295): 223-226.
- Mason, A. R., A. Ziemann and S. Finkbeiner (2014). "Targeting the low-hanging fruit of neurodegeneration." *Neurology* **83**(16): 1470-1473.

- Massman, P. J., J. Sims, N. Cooke, L. J. Haverkamp, V. Appel and S. H. Appel (1996). "Prevalence and correlates of neuropsychological deficits in amyotrophic lateral sclerosis." J Neurol Neurosurg Psychiatry **61**(5): 450-455.
- May, S., D. Hornburg, M. H. Schludi, T. Arzberger, K. Rentzsch, B. M. Schwenk, F. A. Grasser, K. Mori, E. Kremmer, J. Banzhaf-Strathmann, M. Mann, F. Meissner and D. Edbauer (2014). "C9orf72 FTLD/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration." Acta Neuropathol **128**(4): 485-503.
- McHugh, L. and J. W. Arthur (2008). "Computational methods for protein identification from mass spectrometry data." PLoS Comput Biol **4**(2): e12.
- McKhann, G. M., M. S. Albert, M. Grossman, B. Miller, D. Dickson and J. Q. Trojanowski (2001). "Clinical and pathological diagnosis of frontotemporal dementia: report of the Work Group on Frontotemporal Dementia and Pick's Disease." Arch Neurol **58**(11): 1803-1809.
- Mehta, P., V. Antao, W. Kaye, M. Sanchez, D. Williamson, L. Bryan, O. Muravov and K. Horton (2014). "Prevalence of amyotrophic lateral sclerosis - United States, 2010-2011." MMWR Surveill Summ **63 Suppl 7**: 1-14.
- Miller, J. A., M. C. Oldham and D. H. Geschwind (2008). "A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging." J Neurosci **28**(6): 1410-1420.
- Miller, J. A., R. L. Woltjer, J. M. Goodenbour, S. Horvath and D. H. Geschwind (2013). "Genes and pathways underlying regional and cell type changes in Alzheimer's disease." Genome Med **5**(5): 48.
- Miller, Z. A., V. E. Sturm, G. B. Camsari, A. Karydas, J. S. Yokoyama, L. T. Grinberg, A. L. Boxer, H. J. Rosen, K. P. Rankin, M. L. Gorno-Tempini, G. Coppola, D. H. Geschwind, R. Rademakers, W. W. Seeley, N. R. Graff-Radford and B. L. Miller (2016). "Increased prevalence of autoimmune disease within C9 and FTD/MND cohorts: Completing the picture." Neurol Neuroimmunol Neuroinflamm **3**(6): e301.
- Millul, A., E. Beghi, G. Logroscino, A. Micheli, E. Vitelli and A. Zardi (2005). "Survival of patients with amyotrophic lateral sclerosis in a population-based registry." Neuroepidemiology **25**(3): 114-119.
- Mitsuyama, Y. (1993). "Presenile dementia with motor neuron disease." Dementia **4**(3-4): 137-142.
- Mizielinska, S., S. Gronke, T. Niccoli, C. E. Ridler, E. L. Clayton, A. Devoy, T. Moens, F. E. Norona, I. O. Woollacott, J. Pietrzyk, K. Cleverley, A. J. Nicoll, S. Pickering-Brown, J. Dols, M. Cabecinha, O. Hendrich, P. Fratta, E. M. Fisher, L. Partridge and A. M. Isaacs (2014). "C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins." Science **345**(6201): 1192-1194.
- Mizielinska, S., T. Lashley, F. E. Norona, E. L. Clayton, C. E. Ridler, P. Fratta and A. M. Isaacs (2013). "C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci." Acta Neuropathol **126**(6): 845-857.
- Montuschi, A., B. Iazzolino, A. Calvo, C. Moglia, L. Lopiano, G. Restagno, M. Brunetti, I. Ossola, A. Lo Presti, S. Cammarosano, A. Canosa and A. Chio (2015). "Cognitive correlates in

amyotrophic lateral sclerosis: a population-based study in Italy." J Neurol Neurosurg Psychiatry **86**(2): 168-173.

Mori, K., S. Lammich, I. R. Mackenzie, I. Forne, S. Zilow, H. Kretzschmar, D. Edbauer, J. Janssens, G. Kleinberger, M. Cruts, J. Herms, M. Neumann, C. Van Broeckhoven, T. Arzberger and C. Haass (2013). "hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations." Acta Neuropathol **125**(3): 413-423.

Morita, M., A. Al-Chalabi, P. M. Andersen, B. Hosler, P. Sapp, E. Englund, J. E. Mitchell, J. J. Habgood, J. de Belleruche, J. Xi, W. Jongjaroenprasert, H. R. Horvitz, L. G. Gunnarsson and R. H. Brown, Jr. (2006). "A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia." Neurology **66**(6): 839-844.

Murphy, J. M., R. G. Henry, S. Langmore, J. H. Kramer, B. L. Miller and C. Lomen-Hoerth (2007). "Continuum of frontal lobe impairment in amyotrophic lateral sclerosis." Arch Neurol **64**(4): 530-534.

Murray, M. E., M. DeJesus-Hernandez, N. J. Rutherford, M. Baker, R. Duara, N. R. Graff-Radford, Z. K. Wszolek, T. J. Ferman, K. A. Josephs, K. B. Boylan, R. Rademakers and D. W. Dickson (2011). "Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72." Acta Neuropathol **122**(6): 673-690.

Neary, D., J. S. Snowden, L. Gustafson, U. Passant, D. Stuss, S. Black, M. Freedman, A. Kertesz, P. H. Robert, M. Albert, K. Boone, B. L. Miller, J. Cummings and D. F. Benson (1998). "Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria." Neurology **51**(6): 1546-1554.

Neary, D., J. S. Snowden, D. M. Mann, B. Northen, P. J. Goulding and N. Macdermott (1990). "Frontal lobe dementia and motor neuron disease." J Neurol Neurosurg Psychiatry **53**(1): 23-32.

Neumann, M., R. Rademakers, S. Roeber, M. Baker, H. A. Kretzschmar and I. R. A. Mackenzie (2009). "A new subtype of frontotemporal lobar degeneration with FUS pathology." Brain **132**(11): 2922-2931.

Neumann, M., D. M. Sampathu, L. K. Kwong, A. C. Truax, M. C. Micsenyi, T. T. Chou, J. Bruce, T. Schuck, M. Grossman, C. M. Clark, L. F. McCluskey, B. L. Miller, E. Masliah, I. R. Mackenzie, H. Feldman, W. Feiden, H. A. Kretzschmar, J. Q. Trojanowski and V. M. Lee (2006). "Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis." Science **314**(5796): 130-133.

Ng, A. S., R. Rademakers and B. L. Miller (2015). "Frontotemporal dementia: a bridge between dementia and neuromuscular disease." Ann N Y Acad Sci **1338**: 71-93.

Nishimura, A. L., M. Mitne-Neto, H. C. Silva, A. Richieri-Costa, S. Middleton, D. Cascio, F. Kok, J. R. Oliveira, T. Gillingwater, J. Webb, P. Skehel and M. Zatz (2004). "A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis." Am J Hum Genet **75**(5): 822-831.

O'Rourke, J. G., L. Bogdanik, A. Yanez, D. Lall, A. J. Wolf, A. K. Muhammad, R. Ho, S. Carmona, J. P. Vit, J. Zarrow, K. J. Kim, S. Bell, M. B. Harms, T. M. Miller, C. A. Dangler, D. M. Underhill, H. S. Goodridge, C. M. Lutz and R. H. Baloh (2016). "C9orf72 is required for proper macrophage and microglial function in mice." Science **351**(6279): 1324-1329.

Ohki, Y., A. Wenninger-Weinzierl, A. Hruscha, K. Asakawa, K. Kawakami, C. Haass, D. Edbauer and B. Schmid (2017). "Glycine-alanine dipeptide repeat protein contributes to toxicity in a zebrafish model of C9orf72 associated neurodegeneration." Mol Neurodegener **12**(1): 6.

Oldham, M. C. (2014). Transcriptomics: From Differential Expression to Coexpression. The OMICs: Applications in Neuroscience. G. Coppola. New York, Oxford University Press: 85-113.

Oldham, M. C., G. Konopka, K. Iwamoto, P. Langfelder, T. Kato, S. Horvath and D. H. Geschwind (2008). "Functional organization of the transcriptome in human brain." Nat Neurosci **11**(11): 1271-1282.

Parikshak, N. N., M. J. Gandal and D. H. Geschwind (2015). "Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders." Nat Rev Genet **16**(8): 441-458.

Pasinelli, P. and R. H. Brown (2006). "Molecular biology of amyotrophic lateral sclerosis: insights from genetics." Nat Rev Neurosci **7**(9): 710-723.

Polymenidou, M., C. Lagier-Tourenne, K. R. Hutt, S. C. Huelga, J. Moran, T. Y. Liang, S. C. Ling, E. Sun, E. Wancewicz, C. Mazur, H. Kordasiewicz, Y. Sedaghat, J. P. Donohue, L. Shiue, C. F. Bennett, G. W. Yeo and D. W. Cleveland (2011). "Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43." Nat Neurosci **14**(4): 459-468.

Pupillo, E., P. Messina, G. Logroscino and E. Beghi (2014). "Long-term survival in amyotrophic lateral sclerosis: a population-based study." Ann Neurol **75**(2): 287-297.

Radford, R. A., M. Morsch, S. L. Rayner, N. J. Cole, D. L. Pountney and R. S. Chung (2015). "The established and emerging roles of astrocytes and microglia in amyotrophic lateral sclerosis and frontotemporal dementia." Front Cell Neurosci **9**: 414.

Rainero, I., E. Rubino, A. Michelerio, F. D'Agata, S. Gentile and L. Pinessi (2017). "Recent advances in the molecular genetics of frontotemporal lobar degeneration." Funct Neurol **32**(1): 7-16.

Rascovsky, K., J. R. Hodges, D. Knopman, M. F. Mendez, J. H. Kramer, J. Neuhaus, J. C. van Swieten, H. Seelaar, E. G. Dopper, C. U. Onyike, A. E. Hillis, K. A. Josephs, B. F. Boeve, A. Kertesz, W. W. Seeley, K. P. Rankin, J. K. Johnson, M. L. Gorno-Tempini, H. Rosen, C. E. Prioleau-Latham, A. Lee, C. M. Kipps, P. Lillo, O. Piguet, J. D. Rohrer, M. N. Rossor, J. D. Warren, N. C. Fox, D. Galasko, D. P. Salmon, S. E. Black, M. Mesulam, S. Weintraub, B. C. Dickerson, J. Diehl-Schmid, F. Pasquier, V. Deramecourt, F. Lebert, Y. Pijnenburg, T. W. Chow, F. Manes, J. Grafman, S. F. Cappa, M. Freedman, M. Grossman and B. L. Miller (2011). "Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia." Brain **134**(Pt 9): 2456-2477.

Ratnavalli, E., C. Brayne, K. Dawson and J. R. Hodges (2002). "The prevalence of frontotemporal dementia." Neurology **58**(11): 1615-1621.

Ratti, A., L. Corrado, B. Castellotti, R. Del Bo, I. Fogh, C. Cereda, C. Tiloca, C. D'Ascenzo, A. Bagarotti, V. Pensato, M. Ranieri, S. Gagliardi, D. Calini, L. Mazzini, F. Taroni, S. Corti, M. Ceroni, G. D. Oggioni, K. Lin, J. F. Powell, G. Soraru, N. Ticozzi, G. P. Comi, S. D'Alfonso, C.

Gellera and V. Silani (2012). "C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect." *Neurobiol Aging* **33**(10): 2528.e2527-2514.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wu, A. Chio, G. Restagno, G. Borghero, M. Sabatelli, I. Consortium, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R. Morris, P. J. Tienari and B. J. Traynor (2011). "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD." *Neuron* **72**(2): 257-268.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wu, A. Chio, G. Restagno, G. Borghero, M. Sabatelli, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R. Morris, P. J. Tienari and B. J. Traynor (2011). "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD." *Neuron* **72**(2): 257-268.

Roberson, E. D., J. H. Hesse, K. D. Rose, H. Slama, J. K. Johnson, K. Yaffe, M. S. Forman, C. A. Miller, J. Q. Trojanowski, J. H. Kramer and B. L. Miller (2005). "Frontotemporal dementia progresses to death faster than Alzheimer disease." *Neurology* **65**(5): 719-725.

Rosen, D. R., T. Siddique, D. Patterson, D. A. Figlewicz, P. Sapp, A. Hentati, D. Donaldson, J. Goto, J. P. O'Regan, H. X. Deng and et al. (1993). "Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis." *Nature* **362**(6415): 59-62.

Rosso, S. M., E. J. Landweer, M. Houterman, L. Donker Kaat, C. M. van Duijn and J. C. van Swieten (2003). "Medical and environmental risk factors for sporadic frontotemporal dementia: a retrospective case-control study." *J Neurol Neurosurg Psychiatry* **74**(11): 1574-1576.

Rosor, A. M., B. Kalmar, L. Greensmith and M. M. Reilly (2011). "The distal hereditary motor neuropathies." *Journal of Neurology, Neurosurgery & Psychiatry* **83**(1): 6.

Rutherford, N. J., M. G. Heckman, M. DeJesus-Hernandez, M. C. Baker, A. I. Soto-Ortolaza, S. Rayaprolu, H. Stewart, E. Finger, K. Volkening, W. W. Seeley, K. J. Hatanpaa, C. Lomen-Hoerth, A. Kertesz, E. H. Bigio, C. Lippa, D. S. Knopman, H. A. Kretschmar, M. Neumann, R. J. Caselli, C. L. White, 3rd, I. R. Mackenzie, R. C. Petersen, M. J. Strong, B. L. Miller, B. F.

- Boeve, R. J. Uitti, K. B. Boylan, Z. K. Wszolek, N. R. Graff-Radford, D. W. Dickson, O. A. Ross and R. Rademakers (2012). "Length of normal alleles of C9ORF72 GGGGCC repeat do not influence disease phenotype." *Neurobiol Aging* **33**(12): 2950.e2955-2957.
- Sabatelli, M., F. L. Conforti, M. Zollino, G. Mora, M. R. Monsurro, P. Volanti, K. Marinou, F. Salvi, M. Corbo, F. Giannini, S. Battistini, S. Penco, C. Lunetta, A. Quattrone, A. Gambardella, G. Logroscino, I. Simone, I. Bartolomei, F. Pisano, G. Tedeschi, A. Conte, R. Spataro, V. La Bella, C. Caponnetto, G. Mancardi, P. Mandich, P. Sola, J. Mandrioli, A. E. Renton, E. Majounie, Y. Abramzon, F. Marrosu, M. G. Marrosu, M. R. Murru, M. A. Sotgiu, M. Pugliatti, C. Rodolico, I. Consortium, C. Moglia, A. Calvo, I. Ossola, M. Brunetti, B. J. Traynor, G. Borghero, G. Restagno and A. Chio (2012). "C9ORF72 hexanucleotide repeat expansions in the Italian sporadic ALS population." *Neurobiol Aging* **33**(8): 1848 e1815-1820.
- Sampathu, D. M., M. Neumann, L. K. Kwong, T. T. Chou, M. Micsenyi, A. Truax, J. Bruce, M. Grossman, J. Q. Trojanowski and V. M. Lee (2006). "Pathological heterogeneity of frontotemporal lobar degeneration with ubiquitin-positive inclusions delineated by ubiquitin immunohistochemistry and novel monoclonal antibodies." *Am J Pathol* **169**(4): 1343-1352.
- Seelaar, H., W. Kamphorst, S. M. Rosso, A. Azmani, R. Masdjedi, I. de Koning, J. A. Maat-Kievit, B. Anar, L. Donker Kaat, G. J. Breedveld, D. Dooijes, J. M. Rozemuller, I. F. Bronner, P. Rizzu and J. C. van Swieten (2008). "Distinct genetic forms of frontotemporal dementia." *Neurology* **71**(16): 1220-1226.
- Seidel, K., P. Marangoni, C. Tang, B. Houshmand, W. Du, R. L. Maas, S. Murray, M. C. Oldham and O. D. Klein (2017). "Resolving stem and progenitor cells in the adult mouse incisor through gene co-expression analysis." *eLife* **6**: e24712.
- Seyfried, N. T., E. B. Dammer, V. Swarup, D. Nandakumar, D. M. Duong, L. Yin, Q. Deng, T. Nguyen, C. M. Hales, T. Wingo, J. Glass, M. Gearing, M. Thambisetty, J. C. Troncoso, D. H. Geschwind, J. J. Lah and A. I. Levey (2017). "A Multi-network Approach Identifies Protein-Specific Co-expression in Asymptomatic and Symptomatic Alzheimer's Disease." *Cell Syst* **4**(1): 60-72.e64.
- Seyfried, N. T., Y. M. Gozal, L. E. Donovan, J. H. Herskowitz, E. B. Dammer, Q. Xia, L. Ku, J. Chang, D. M. Duong, H. D. Rees, D. S. Cooper, J. D. Glass, M. Gearing, M. G. Tansey, J. J. Lah, Y. Feng, A. I. Levey and J. Peng (2012). "Quantitative analysis of the detergent-insoluble brain proteome in frontotemporal lobar degeneration using SILAC internal standards." *J Proteome Res* **11**(5): 2721-2738.
- Simon-Sanchez, J., E. G. Dopper, P. E. Cohn-Hokke, R. K. Hukema, N. Nicolaou, H. Seelaar, J. R. de Graaf, I. de Koning, N. M. van Schoor, D. J. Deeg, M. Smits, J. Raaphorst, L. H. van den Berg, H. J. Schelhaas, C. E. De Die-Smulders, D. Majoor-Krakauer, A. J. Rozemuller, R. Willemsen, Y. A. Pijnenburg, P. Heutink and J. C. van Swieten (2012). "The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions." *Brain* **135**(Pt 3): 723-735.
- Snowden, J. S., J. Harris, A. Richardson, S. Rollinson, J. C. Thompson, D. Neary, D. M. Mann and S. Pickering-Brown (2013). "Frontotemporal dementia with amyotrophic lateral sclerosis: a clinical comparison of patients with and without repeat expansions in C9orf72." *Amyotroph Lateral Scler Frontotemporal Degener* **14**(3): 172-176.

Snowden, J. S., S. Rollinson, J. C. Thompson, J. M. Harris, C. L. Stopford, A. M. Richardson, M. Jones, A. Gerhard, Y. S. Davidson, A. Robinson, L. Gibbons, Q. Hu, D. DuPlessis, D. Neary, D. M. Mann and S. M. Pickering-Brown (2012). "Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations." *Brain* **135**(Pt 3): 693-708.

Sreedharan, J., I. P. Blair, V. B. Tripathi, X. Hu, C. Vance, B. Rogelj, S. Ackerley, J. C. Durnall, K. L. Williams, E. Buratti, F. Baralle, J. de Belleruche, J. D. Mitchell, P. N. Leigh, A. Al-Chalabi, C. C. Miller, G. Nicholson and C. E. Shaw (2008). "TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis." *Science* **319**(5870): 1668-1672.

Stalekar, M., X. Yin, K. Rebolj, S. Darovic, C. Troakes, M. Mayr, C. E. Shaw and B. Rogelj (2015). "Proteomic analyses reveal that loss of TDP-43 affects RNA processing and intracellular transport." *Neuroscience* **293**: 157-170.

Stewart, H., N. J. Rutherford, H. Briemberg, C. Krieger, N. Cashman, M. Fabros, M. Baker, A. Fok, M. DeJesus-Hernandez, A. Eisen, R. Rademakers and I. R. Mackenzie (2012). "Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p." *Acta Neuropathol* **123**(3): 409-417.

Sudria-Lopez, E., M. Koppers, M. de Wit, C. van der Meer, H. J. Westeneng, C. A. Zundel, S. A. Youssef, L. Harkema, A. de Bruin, J. H. Veldink, L. H. van den Berg and R. J. Pasterkamp (2016). "Full ablation of C9orf72 in mice causes immune system-related pathology and neoplastic events but no motor neuron defects." *Acta Neuropathol* **132**(1): 145-147.

Suh, E., E. B. Lee, D. Neal, E. M. Wood, J. B. Toledo, L. Rennert, D. J. Irwin, C. T. McMillan, B. Krock, L. B. Elman, L. F. McCluskey, M. Grossman, S. X. Xie, J. Q. Trojanowski and V. M. Van Deerlin (2015). "Semi-automated quantification of C9orf72 expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration." *Acta Neuropathol* **130**(3): 363-372.

Sullivan, P. M., X. Zhou, A. M. Robins, D. H. Paushter, D. Kim, M. B. Smolka and F. Hu (2016). "The ALS/FTLD associated protein C9orf72 associates with SMCR8 and WDR41 to regulate the autophagy-lysosome pathway." *Acta Neuropathol Commun* **4**(1): 51.

Taylor, J. P., R. H. Brown, Jr. and D. W. Cleveland (2016). "Decoding ALS: from genes to mechanism." *Nature* **539**(7628): 197-206.

Tran, H., S. Almeida, J. Moore, T. F. Gendron, U. Chalasani, Y. Lu, X. Du, J. A. Nickerson, L. Petrucelli, Z. Weng and F. B. Gao (2015). "Differential Toxicity of Nuclear RNA Foci versus Dipeptide Repeat Proteins in a Drosophila Model of C9ORF72 FTD/ALS." *Neuron* **87**(6): 1207-1214.

Traxinger, K., C. Kelly, B. A. Johnson, R. H. Lyles and J. D. Glass (2013). "Prognosis and epidemiology of amyotrophic lateral sclerosis: Analysis of a clinic population, 1997-2011." *Neurol Clin Pract* **3**(4): 313-320.

Troakes, C., S. Maekawa, L. Wijesekera, B. Rogelj, L. Siklos, C. Bell, B. Smith, S. Newhouse, C. Vance, L. Johnson, T. Hortobagyi, A. Shatunov, A. Al-Chalabi, N. Leigh, C. E. Shaw, A. King and S. Al-Sarraj (2012). "An MND/ALS phenotype associated with C9orf72 repeat expansion: abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline." *Neuropathology* **32**(5): 505-514.

Tsuji, H., T. Arai, F. Kametani, T. Nonaka, M. Yamashita, M. Suzukake, M. Hosokawa, M. Yoshida, H. Hatsuta, M. Takao, Y. Saito, S. Murayama, H. Akiyama, M. Hasegawa, D. M. Mann and A. Tamaoka (2012). "Molecular analysis and biochemical classification of TDP-43 proteinopathy." *Brain* **135**(Pt 11): 3380-3391.

Tyanova, S., T. Temu, P. Sinitcyn, A. Carlson, M. Y. Hein, T. Geiger, M. Mann and J. Cox (2016). "The Perseus computational platform for comprehensive analysis of (prote)omics data." *Nat Methods* **13**(9): 731-740.

Umoh, M. E., C. Fournier, Y. Li, M. Polak, L. Shaw, J. E. Landers, W. Hu, M. Gearing and J. D. Glass (2016). "Comparative analysis of C9orf72 and sporadic disease in an ALS clinic population." *Neurology*.

Valdmanis, P. N., N. Dupre, J. P. Bouchard, W. Camu, F. Salachas, V. Meininger, M. Strong and G. A. Rouleau (2007). "Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p." *Arch Neurol* **64**(2): 240-245.

Valle, J., E. Roberts, S. Paulukonis, N. Collins, P. English and W. Kaye (2015). "Epidemiology and surveillance of amyotrophic lateral sclerosis in two large metropolitan areas in California." *Amyotroph Lateral Scler Frontotemporal Degener* **16**(3-4): 209-215.

van Blitterswijk, M., M. DeJesus-Hernandez, E. Niemantsverdriet, M. E. Murray, M. G. Heckman, N. N. Diehl, P. H. Brown, M. C. Baker, N. A. Finch, P. O. Bauer, G. Serrano, T. G. Beach, K. A. Josephs, D. S. Knopman, R. C. Petersen, B. F. Boeve, N. R. Graff-Radford, K. B. Boylan, L. Petrucelli, D. W. Dickson and R. Rademakers (2013). "Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study." *Lancet Neurol* **12**(10): 978-988.

van Blitterswijk, M., M. DeJesus-Hernandez and R. Rademakers (2012). "How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders?" *Curr Opin Neurol* **25**(6): 689-700.

van Blitterswijk, M., B. Mullen, M. G. Heckman, M. C. Baker, M. DeJesus-Hernandez, P. H. Brown, M. E. Murray, G. Y. Hsiung, H. Stewart, A. M. Karydas, E. Finger, A. Kertesz, E. H. Bigio, S. Weintraub, M. Mesulam, K. J. Hatanpaa, C. L. White, 3rd, M. Neumann, M. J. Strong, T. G. Beach, Z. K. Wszolek, C. Lippa, R. Caselli, L. Petrucelli, K. A. Josephs, J. E. Parisi, D. S. Knopman, R. C. Petersen, I. R. Mackenzie, W. W. Seeley, L. T. Grinberg, B. L. Miller, K. B. Boylan, N. R. Graff-Radford, B. F. Boeve, D. W. Dickson and R. Rademakers (2014). "Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers." *Neurobiol Aging* **35**(10): 2421.e2413-2427.

van der Zee, J., I. Gijssels, L. Dillen, T. Van Langenhove, J. Theuns, S. Engelborghs, S. Philtjens, M. Vandenbulcke, K. Sleegers, A. Sieben, V. Baumer, G. Maes, E. Corsmit, B. Borroni, A. Padovani, S. Archetti, R. Perneczky, J. Diehl-Schmid, A. de Mendonca, G. Miltenberger-Miltenyi, S. Pereira, J. Pimentel, B. Nacmias, S. Bagnoli, S. Sorbi, C. Graff, H. H. Chiang, M. Westerlund, R. Sanchez-Valle, A. Llado, E. Gelpi, I. Santana, M. R. Almeida, B. Santiago, G. Frisoni, O. Zanetti, C. Bonvicini, M. Synofzik, W. Maetzler, J. M. Vom Hagen, L. Schols, M. T. Heneka, F. Jessen, R. Matej, E. Parobkova, G. G. Kovacs, T. Strobel, S. Sarafov, I. Tournev, A. Jordanova, A. Danek, T. Arzberger, G. M. Fabrizi, S. Testi, E. Salmon, P. Santens, J. J. Martin, P. Cras, R. Vandenberghe, P. P. De Deyn, M. Cruts, C. Van Broeckhoven, J. van der Zee, I. Gijssels, L. Dillen, T. Van Langenhove, J. Theuns, S. Philtjens, K. Sleegers, V.

Baumer, G. Maes, E. Corsmit, M. Cruts, C. Van Broeckhoven, J. van der Zee, I. Gijselinck, L. Dillen, T. Van Langenhove, S. Philtjens, J. Theuns, K. Sleegers, V. Baumer, G. Maes, M. Cruts, C. Van Broeckhoven, S. Engelborghs, P. P. De Deyn, P. Cras, S. Engelborghs, P. P. De Deyn, M. Vandenbulcke, M. Vandenbulcke, B. Borroni, A. Padovani, S. Archetti, R. Perneczky, J. Diehl-Schmid, M. Synofzik, W. Maetzler, J. Muller Vom Hagen, L. Schols, M. Synofzik, W. Maetzler, J. Muller Vom Hagen, L. Schols, M. T. Heneka, F. Jessen, A. Ramirez, D. Kurzweily, C. Sachtleben, W. Mairer, A. de Mendonca, G. Miltenberger-Miltenyi, S. Pereira, C. Firmo, J. Pimentel, R. Sanchez-Valle, A. Llado, A. Antonell, J. Molinuevo, E. Gelpi, C. Graff, H. H. Chiang, M. Westerlund, C. Graff, A. Kinhult Stahlbom, H. Thonberg, I. Nennesmo, A. Borjesson-Hanson, B. Nacmias, S. Bagnoli, S. Sorbi, V. Bessi, I. Piaceri, I. Santana, B. Santiago, I. Santana, M. Helena Ribeiro, M. Rosario Almeida, C. Oliveira, J. Massano, C. Garret, P. Pires, G. Frisoni, O. Zanetti, C. Bonvicini, S. Sarafov, I. Tournev, A. Jordanova, I. Tournev, G. G. Kovacs, T. Strobel, M. T. Heneka, F. Jessen, A. Ramirez, D. Kurzweily, C. Sachtleben, W. Mairer, F. Jessen, R. Matej, E. Parobkova, A. Danel, T. Arzberger, G. Maria Fabrizi, S. Testi, S. Ferrari, T. Cavallaro, E. Salmon, P. Santens and P. Cras (2013). "A pan-European study of the C9orf72 repeat associated with FTLN: geographic prevalence, genomic instability, and intermediate repeats." *Hum Mutat* **34**(2): 363-373.

Van Mossevelde, S., J. van der Zee, M. Cruts and C. Van Broeckhoven (2017). "Relationship between C9orf72 repeat size and clinical phenotype." *Curr Opin Genet Dev* **44**: 117-124.

van Rheenen, W., M. van Blitterswijk, M. H. Huisman, L. Vlam, P. T. van Doormaal, M. Seelen, J. Medic, D. Dooijes, M. de Visser, A. J. van der Kooij, J. Raaphorst, H. J. Schelhaas, W. L. van der Pol, J. H. Veldink and L. H. van den Berg (2012). "Hexanucleotide repeat expansions in C9ORF72 in the spectrum of motor neuron diseases." *Neurology* **79**(9): 878-882.

van Swieten, J. C. and P. Heutink (2008). "Mutations in progranulin (GRN) within the spectrum of clinical and pathological phenotypes of frontotemporal dementia." *Lancet Neurol* **7**(10): 965-974.

Vance, C., A. Al-Chalabi, D. Ruddy, B. N. Smith, X. Hu, J. Sreedharan, T. Siddique, H. J. Schelhaas, B. Kusters, D. Troost, F. Baas, V. de Jong and C. E. Shaw (2006). "Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3." *Brain* **129**(Pt 4): 868-876.

Voineagu, I., X. Wang, P. Johnston, J. K. Lowe, Y. Tian, S. Horvath, J. Mill, R. M. Cantor, B. J. Blencowe and D. H. Geschwind (2011). "Transcriptomic analysis of autistic brain reveals convergent molecular pathology." *Nature* **474**(7351): 380-384.

Waite, A. J., D. Baumer, S. East, J. Neal, H. R. Morris, O. Ansorge and D. J. Blake (2014). "Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion." *Neurobiol Aging* **35**(7): 1779.e1775-1779.e1713.

Webster, C. P., E. F. Smith, C. S. Bauer, A. Moller, G. M. Hautbergue, L. Ferraiuolo, M. A. Myszczyńska, A. Higginbottom, M. J. Walsh, A. J. Whitworth, B. K. Kaspar, K. Meyer, P. J. Shaw, A. J. Grierson and K. J. De Vos (2016). "The C9orf72 protein interacts with Rab1a and the ULK1 complex to regulate initiation of autophagy." *The EMBO Journal* **35**(15): 1656-1676.

Weder, N. D., R. Aziz, K. Wilkins and R. R. Tampi (2007). "Frontotemporal dementias: a review." *Ann Gen Psychiatry* **6**: 15.

- Wen, X., W. Tan, T. Westergard, K. Krishnamurthy, S. S. Markandiah, Y. Shi, S. Lin, N. A. Shneider, J. Monaghan, U. B. Pandey, P. Pasinelli, J. K. Ichida and D. Trotti (2014). "Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate in vitro and in vivo neuronal death." Neuron **84**(6): 1213-1225.
- Williams, J. R., D. Fitzhenry, L. Grant, D. Martyn and D. A. Kerr (2013). "Diagnosis pathway for patients with amyotrophic lateral sclerosis: retrospective analysis of the US Medicare longitudinal claims database." BMC Neurology **13**: 160-160.
- Wingo, T. S., D. J. Cutler, N. Yarab, C. M. Kelly and J. D. Glass (2011). "The heritability of amyotrophic lateral sclerosis in a clinically ascertained United States research registry." PLoS One **6**(11): e27985.
- Wu, Y. E., N. N. Parikshak, T. G. Belgard and D. H. Geschwind (2016). "Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder." Nat Neurosci **19**(11): 1463-1476.
- Xiao, S., L. MacNair, P. McGoldrick, P. M. McKeever, J. R. McLean, M. Zhang, J. Keith, L. Zinman, E. Rogaeva and J. Robertson (2015). "Isoform-specific antibodies reveal distinct subcellular localizations of C9orf72 in amyotrophic lateral sclerosis." Ann Neurol **78**(4): 568-583.
- Xu, P., D. M. Duong and J. Peng (2009). "Systematical Optimization of Reverse-Phase Chromatography for Shotgun Proteomics." Journal of Proteome Research **8**(8): 3944-3950.
- Xu, Z., M. Poidevin, X. Li, Y. Li, L. Shu, D. L. Nelson, H. Li, C. M. Hales, M. Gearing, T. S. Wingo and P. Jin (2013). "Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration." Proc Natl Acad Sci U S A **110**(19): 7778-7783.
- Xu, Z. S. (2012). "Does a loss of TDP-43 function cause neurodegeneration?" Mol Neurodegener **7**: 27.
- Yang, M., C. Liang, K. Swaminathan, S. Herrlinger, F. Lai, R. Shiekhattar and J. F. Chen (2016). "A C9ORF72/SMCR8-containing complex regulates ULK1 and plays a dual role in autophagy." Sci Adv **2**(9): e1601167.
- Yip, A. M. and S. Horvath (2007). "Gene network interconnectedness and the generalized topological overlap measure." BMC Bioinformatics **8**(1): 22.
- Zhang, B. and S. Horvath (2005). "A general framework for weighted gene co-expression network analysis." Stat Appl Genet Mol Biol **4**: Article17.
- Zhang, D., L. M. Iyer, F. He and L. Aravind (2012). "Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease." Front Genet **3**: 283.
- Zhang, K., C. J. Donnelly, A. R. Haeusler, J. C. Grima, J. B. Machamer, P. Steinwald, E. L. Daley, S. J. Miller, K. M. Cunningham, S. Vidensky, S. Gupta, M. A. Thomas, I. Hong, S. L. Chiu, R. L. Haganir, L. W. Ostrow, M. J. Matunis, J. Wang, R. Sattler, T. E. Lloyd and J. D. Rothstein (2015). "The C9orf72 repeat expansion disrupts nucleocytoplasmic transport." Nature **525**(7567): 56-61.

Zhang, Y., K. Chen, S. A. Sloan, M. L. Bennett, A. R. Scholze, S. O'Keefe, H. P. Phatnani, P. Guarnieri, C. Caneda, N. Ruderisch, S. Deng, S. A. Liddelow, C. Zhang, R. Daneman, T. Maniatis, B. A. Barres and J. Q. Wu (2014). "An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex." J Neurosci **34**(36): 11929-11947.

Zhang, Y. J., T. F. Gendron, J. C. Grima, H. Sasaguri, K. Jansen-West, Y. F. Xu, R. B. Katzman, J. Gass, M. E. Murray, M. Shinohara, W. L. Lin, A. Garrett, J. N. Stankowski, L. Daugherty, J. Tong, E. A. Perkerson, M. Yue, J. Chew, M. Castanedes-Casey, A. Kurti, Z. S. Wang, A. M. Liesinger, J. D. Baker, J. Jiang, C. Lagier-Tourenne, D. Edbauer, D. W. Cleveland, R. Rademakers, K. B. Boylan, G. Bu, C. D. Link, C. A. Dickey, J. D. Rothstein, D. W. Dickson, J. D. Fryer and L. Petrucelli (2016). "C9ORF72 poly(GA) aggregates sequester and impair HR23 and nucleocytoplasmic transport proteins." Nat Neurosci **19**(5): 668-677.

Zhang, Y. J., Y. F. Xu, C. Cook, T. F. Gendron, P. Roettges, C. D. Link, W. L. Lin, J. Tong, M. Castanedes-Casey, P. Ash, J. Gass, V. Rangachari, E. Buratti, F. Baralle, T. E. Golde, D. W. Dickson and L. Petrucelli (2009). "Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity." Proc Natl Acad Sci U S A **106**(18): 7607-7612.

Ziegler, L. H. (1930). "PSychotic and emotional phenomena associated with amyotrophic lateral sclerosis." Archives of Neurology & Psychiatry **24**(5): 930-936.

Zu, T., Y. Liu, M. Banez-Coronel, T. Reid, O. Pletnikova, J. Lewis, T. M. Miller, M. B. Harms, A. E. Falchook, S. H. Subramony, L. W. Ostrow, J. D. Rothstein, J. C. Troncoso and L. P. Ranum (2013). "RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia." Proc Natl Acad Sci U S A **110**(51): E4968-4977.

#### Funding sources:

*Pre-doctoral T32 Neurology training Grant (3T32NS007480-16), Emory Neurosciences NINDS core facilities P30 grant (NS055077), Alzheimer's Disease Research Center NIH-NIA P50 grant (AG025688)*