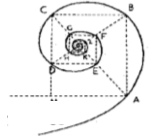


UNIVERSITÀ DEGLI STUDI DI MILANO

SCUOLA DI DOTTORATO IN MEDICINA MOLECOLARE

CICLO XXVII

Anno Accademico 2013/2014

TESI DI DOTTORATO DI RICERCA

Settore Scientifico Disciplinare MED/26

EXOME SEQUENCING APPROACH TO IDENTIFY CAUSATIVE GENES FOR AMYOTROPHIC LATERAL SCLEROSIS

Dottorando : Cinzia Tiloca

Matricola N° R09756

TUTORE: Ch.mo Prof. Giacomo Pietro Comi

CO-TUTORE: Ch.mo Prof. Vincenzo Silani

COORDINATORE DEL DOTTORATO: Ch.mo Prof. Mario Clerici

ABSTRACT

Introduction Amyotrophic lateral sclerosis (ALS) is an adult-onset neurodegenerative disorder caused by the loss of motor neurons in the cerebral cortex, brainstem and spinal cord. ALS occurs prevalently as sporadic forms (SALS), but a small proportion of cases (5-10%) displays a positive family history (FALS), generally with an autosomal dominant pattern of inheritance. To date, more than 20 causative genes have been identified in FALS, providing fundamental insights into the pathogenic mechanisms and underlying the great genetic heterogeneity of the disease. Despite these numerous advances, the genetic basis of nearly 40% of FALS remains to be identified, while the genetic component of SALS is largely unknown. A powerful and innovative tool for genetic studies in ALS is represented by next-generation sequencing and in particular by the targeted sequencing of the coding part of the genome or exome.

Aim of this research project was to identify novel genes associated to FALS and SALS by applying complementary approaches all based on exome sequencing, which overcomes the limitations of traditional genetic strategies.

Methods Three different disease gene identification strategies were applied: I) exome sequencing associated to linkage analysis in two large ALS dominant pedigrees; II) exome-wide rare variant burden analysis on 363 unrelated index FALS cases; III) exome sequencing of 32 SALS and their unaffected parents (trio-design).

Results I) By performing exome-sequencing in combination with linkage analysis, we identified *PFN1 (*profilin-1)*,* encoding for a protein regulating actin dynamics, as a novel ALS-causative gene. Mutations in *PFN1* were observed in ~2.6% of FALS and functional studies demonstrated

aggregation propensity, reduction of actin binding ability and axonal outgrowth inhibition of mutant PFN1 proteins.

II) As a result of the unbiased case-control rare variant analysis, applied on a total 12.495 genes, we identified *TUBA4A (*Tubulin, Alpha 4a*)* as candidate gene showing a statistically significant excess of rare damaging variants in 363 index FALS cases sequenced. Functional analysis revealed that ALS-related mutants were defective in forming alpha/beta tubulin dimers *in vitro* and in incorporating into microtubules *in vivo*. In addition, the truncated mutant TUBA4A p.W407X showed aggregation propensities.

III) By sequencing the exomes of 32 SALS patients and their unaffected parents, we identified 25 *de novo* mutations (DNMs) in 16 of 32 trios, with an overall DNM rate of 0.78. Although we did not find recurrently mutated genes in our ALS trios, bionformatic analysis showed potential inter-connections between the candidate genes. Functional classification revealed that DNMs are enriched in genes encoding for proteins involved in transport and in GTPase regulatory activity.

Conclusions Our findings indicate that exome-sequencing, combined with different strategies for study design and data analysis, is an effective and successful approach for the identification of novel ALS causative genes. The identification of *PFN1* and *TUBA4A* genes, encoding for proteins implicated, respectively, in actin polymerization and microtubule formation, further supports a major role of cytoskeletal defects in ALS pathogenesis.

SOMMARIO

Introduzione La Sclerosi laterale amiotrofica (SLA) è una malattia neurodegenerativa progressiva e fatale caratterizzata dalla perdita selettiva dei motoneuroni nella corteccia cerebrale, nel tronco cerebrale e midollo spinale. La maggior parte dei casi è costituita da forme sporadiche, mentre solo il 5-10% dei casi è rappresentato da forme familiari, causate da geni con modalità di trasmissione mendeliana, generalmente autosomica dominante. Sono stati identificati più di 20 geni causativi delle forme familiari, che hanno contribuito a comprendere meglio i meccanismi patogenetici coinvolti e sottolineano la grande eterogeneità genetica della malattia. Nonostante i numerosi progressi raggiunti, in circa il 40% dei casi familiari la causa genetica non è stata ancora identificata, mentre la componente genetica delle forme sporadiche è in gran parte sconosciuta.
L' applicazione delle tecniche di sequenziamento di nuova generazione ed in particolare il sequenziamento della porzione codificante del genoma o esoma rappresenta un approccio innovativo e promettente per gli studi genetici sulla SLA. Lo scopo del presente progetto di ricerca di Dottorato è stato quello di identificare nuovi geni associati alle forme familiari e sporadiche di SLA mediante sequenziamento dell'esoma come metodo alternativo per superare i limiti delle tecniche genetiche tradizionali.

Metodi Sono state utilizzate tre diverse strategie per l'identificazione di geni causativi: I) sequenziamento dell'esoma in combinazione con analisi di linkage in due grandi famiglie SLA a trasmissione dominante; II) analisi per varianti rare degli esomi di 363 casi familiari singoli; III) sequenziamento dell'esoma in 32 casi di SLA sporadica e dei loro genitori non affetti (approccio dei trios).

Risultati I) Attraverso l'approccio combinato di sequenziamento dell'esoma ed analisi di linkage, abbiamo identificato il gene *PFN1*

(profilina-1), codificante per una proteina implicata nella regolazione dell'actina, come nuovo gene causativo di SLA. Mutazioni a carico del gene *PFN1* sono state osservate nel 2,6% dei pazienti SLA familiari e gli studi funzionali condotti sui mutanti hanno dimostrato una maggiore tendenza all'aggregazione, una riduzione del legame all'actina ed un effetto inibitorio sulla crescita assonale.

II) L'analisi delle varianti rare tra casi e controlli, applicata su un totale di 12.495 geni, ha portato all'identificazione di *TUBA4A* (codificante per lalfa-tubulina 4a) come gene candidato caratterizzato da un eccesso significativo di varianti rare potenzialmente dannose nei 363 casi familiari analizzati. L'analisi funzionale ha dimostrato per i mutanti di TUBA4A una capacità ridotta di dimerizzazione con la beta-tubulina *in vitro* ed un'alterata incorporazione nei microtubuli *in vivo*. Inoltre, il mutante tronco TUBA4A p.W407X ha mostrato una maggiore tendenza all'aggregazione.

III) Infine, analizzando l'esoma di 32 pazienti con SLA sporadica e dei loro genitori non affetti, abbiamo identificato 25 mutazioni *de novo* in 16 dei 32 trios analizzati, con un tasso di mutazioni de novo pari a 0,78. Non sono stati identificati geni con molteplici mutazioni *de novo* nei trios sequenziati, ma le analisi bioinformatiche hanno mostrato possibili connessioni tra i geni candidati e la classificazione funzionale ha rilevato che le mutazioni *de novo* sono principalmente a carico di geni codificanti per trasportatori o per proteine con attività regolatoria sulle GTPasi.

Conclusioni I risultati ottenuti hanno dimostrato che il sequenziamento dell'esoma, applicato con specifiche strategie di studio e di analisi, è un approccio efficace per l'identificazione di nuovi geni causativi nella SLA. La scoperta dei due geni *PFN1* e *TUBA4A*, codificanti per proteine coinvolte nel processo di polimerizzazione dell'actina e dei microtubuli, fornisce ulteriori evidenze a supporto del coinvolgimento del citoscheletro nella patogenesi della SLA.

TABLE OF CONTENTS

1. INTRODUCTION 1

1.1 Amyotrophic Lateral Sclerosis (ALS) 1

1.1.1 Epidemiological and clinical features 3

1.1.2 Clinical diagnosis 8

1.1.3 Etiology of ALS 10

1.1.4 Role of enviromental risk factors 11

1.2 Genetics of ALS 14

1.2.1 Causative genes 14

1.2.2 Genetic risk factors 20

1.2.3 Methodologies used to identify ALS genes 22

1.2.3.1 Linkage analysis 22

1.2.3.2 Candidate gene approach 23

1.2.3.3 Genome-wide association studies 24

1.2.3.4 Whole-exome sequencing 26

1.2.4 Problems and limitations in the genetic study of ALS 28

1.3 Pathogenic mechanisms 30

1.3.1 Protein aggregation 30

1.3.2 RNA metabolism 31

1.3.3 Mitochondrial dysfunction 32

1.3.4 Cytoskeletal abnormalities 32

2. AIM OF THE PROJECT 34

3. METHODS AND MATERIALS 35

3.1 Patients and controls 35

3.2 Linkage analysis and exome sequencing 39

3.3 Data analysis strategies 40

3.4 Bioinformatic analysis 42

3.5 Sanger sequencing and genotyping 43

3.6 Cell cultures 45

3.7 Immunofluorescence 46

3.8 Soluble/insoluble fractionation 47

3.9 Immunoprecipitation 48

3.10 Tubulin dimerization assay 48

4. RESULTS 49

4.2 Exome sequencing of ALS families identifies mutations in *PFN1* gene 49

4.2 PFN1 mutants are aggregation prone and show alterations in actin binding 55

4.3 Exome-sequencing of unrelated FALS indentifies *TUBA4A* as candidate gene **61**

4.4 TUBA4A mutants affect microtubule polymerization **66**

4.5 Identification of *de novo* mutations using a trio-based exome sequencing approach **72**

5. DISCUSSION **79**

6. CONCLUSIONS **85**

7. REFERENCES **87**

8. ACKNOWLEDGMENTS **100**

9. SCIENTIFIC PRODUCTION **101**

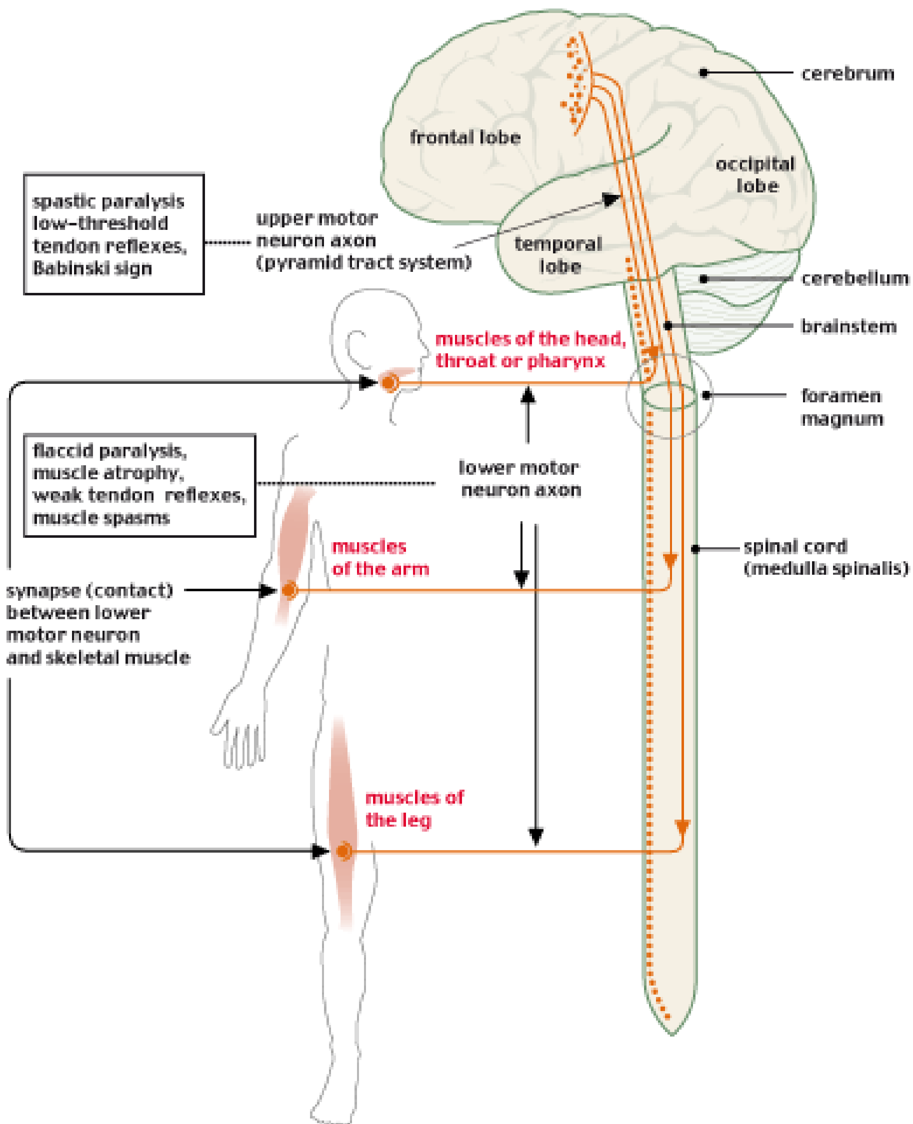
1. INTRODUCTION

1.1 Amyotrophic Lateral Sclerosis (ALS)

Amyotrophic lateral sclerosis (ALS) is a devastating and fatal adult-onset neurodegenerative disorder characterized by the loss of motor neurons in the cerebral cortex, brainstem and spinal cord. The main clinical features of the disease include progressive weakness, wasting and paralysis of skeletal muscles [1]. In ALS both upper motor neurons (UMNs) and lower motor neurons (LMNs) are affected: UMNs cell bodies are located within the primary motor cortex (precentral gyrus) and project their long axons to the lower motor neurons via the corticospinal and corticopontine/bulbar tracts; LMNs are located within the motor nuclei of cranial nerves in the brainstem and the anterior horn of the spinal cord, and directly innervate skeletal muscles (Figure 1). The clinical symptoms depend on the different involvement of UMNs and LMNs. Symptoms of UMNs involvement include spasticity, hyperreflexia and the presence of abnormal reflexes (Babinski's sign). Symptoms of LMNs degeneration include muscle weakness and atrophy, muscle cramps, and fasciculations [2]. The term “Amyotrophic Lateral Sclerosis” was originally introduced in 1874 by the famous French neurologist Jean-Marie Charcot, who was the first to accurately provide a complete clinical description of the disease, correlating the clinical signs with anatomo-pathological findings [3]. “Amyotrophic” refers to the atrophy of muscle fibers, and “lateral sclerosis” describes the specific sclerotic changes in the lateral tracts of the spinal cord due to the degeneration of the corticospinal tracts with reactive astrogliosis.

Considering the great advancements of Jean-Marie Charcot in describing its clinical and pathological features, ALS is still referred to as Charcot’s disease in many parts of the world [4].

Figure 1. Anatomical organization of the human motor systems. The upper motor neurons (UMN), originating from the motor cortex, form synapses with the lower motor neurones (LMN) in the motor nuclei of the brainstem and spinal cord.



1.1.1 Epidemiological and clinical features

ALS is an adult-onset disease with an incidence of 1-2 /100.000 individuals per year in Caucasian populations and a lifetime risk of 1:400. Incidence rate increases with age, reaching a peak between 60 and 80 years, and is higher in male than female (male:female ratio of 1.5) [1].

A lower incidence rate has been reported in some ethnic populations (American Indians and Alaska natives), while it is 50-100 times higher in three geographical areas of the Western Pacific Ocean (Guam, Western New Guinea and Japan's Kii Peninsula) [5]. In these geographical regions the extremely high incidence of ALS, which often occur in combination with parkinsonism and dementia (ALS-PDC), seems to be caused by environmental factors, such as dietary exposure to β-methylamino-L-alanine (BMAA) and high levels of mineral content in the soils and drinking water.

The disease is clinically heterogeneous with an extremely variable clinic presentation as regards age of onset, disease duration, site of onset, relative mix of UMN and LMN deficits and other clinical features. The mean age of onset is between 55-65 years with a median age of onset of 64 years [6]. Only 5% of patients have symptom onset before 25 years of age: these cases are known as juvenile ALS (JALS) and usually present a slower disease progression and a mainly UMN phenotype. In ALS the death occurs from respiratory failure within 2-5 years from symptom onset (with a median survival of 3 years), but disease duration varies widely in individual patients, ranging from few months to over 10 years [7]. To date there is no effective treatment for the disease: the only drug approved by FDA is riluzole, which has been shown to increase survival by about 3 months.

Approximately 75-85% of ALS patients have a classical spinal onset of the disease, with a focal and unilateral muscle weakness starting distally or

proximally in the upper or lower limbs [8]. The frequency of upper limb versus lower limb involvement is approximately equal. Early signs include foot drop, difficulty in walking, loss of hand dexterity or weakness when lifting the arms. As disease progresses and limb function deteriorates, patients lose the ability of walking independently and become dependent on caregivers. About 25% of ALS cases show a bulbar onset, characterized by dysarthria, hypophonia with nasal voice, tongue fasciculations and dysphagia initially for liquids and subsequently for solids [8]. In bulbar patients, limb symptoms can appear almost simultaneously or can occur later, usually within 1-2 years. Bulbar onset, which is associated with a shorter survival, is more frequent in women and in older age groups: 43% of patients over the age of 70 present with bulbar symptoms compared to 15% below the age of 30 [8]. A smaller proportion of individuals (~5%) experiences "respiratory onset" ALS, with initial involvement of the diaphragm and intercostal muscles.

Other characteristic phenotypic variants of ALS, with a typical slower progression, include: the “flail arm” syndrome, which involves predominant muscle weakness in the proximal upper extremities and a relative sparing of lower limbs, and the “flail leg” syndrome or pseudopolyneuritic variant, which is characterized by symmetrical LMN signs within the distal segments of the lower extremities [9].

Other related forms of motor neuron disease include: a) primary lateral sclerosis (PLS) characterized by the exclusive affection of the UMNs; b) PMA (progressive muscular atrophy) resulting from the exclusive affection of LMNs; c) progressive bulbar palsy (PBP) in which the lowest motor neurons of the brain stem are most affected. Long term follow-up of PLS and PMA cases has shown that patients, after a disease duration of six years or more, may progress to ALS, suggesting that these conditions should be considered variants of ALS [10].

Symptoms often begin focally but with progression of disease they spread along continuous anatomic paths. Notably, two groups of motor neurons are relatively preserved until the late stages of the disease: the oculomotor nuclei in the brainstem and the Onuf's nucleus in the sacral spinal cord, which innervates the external striate urethral and rectal sphincters [11]. Recently, researchers are investigating the difference in properties of these neuronal populations, to acquire novel insights into the mechanisms of neuronal degeneration, and to identify potential targets for therapeutic manipulation [12].

Since the disease primarily involves motor neurons, sensory functions are generally preserved, although a minority of patients complains of some numbness and paresthesias. Abnormalities of sensory nerve conduction have been reported in a small number of patients with ALS but these findings often reflect the presence of an unrelated, coexistent condition [8].

Association with other clinical conditions

The concept of ALS as a purely motor neuron disease has been surpassed by the increasing recognition that the disease may be associated with variable clinical conditions, such as cognitive impairment, FTD, aphasia, or parkinsonism [13].

Cognitive impairment in ALS was considered uncommon until recently, because it is often underreported by families and difficult for clinician to assess due to the presence of motor symptoms. In Charcot's original reports cognitive changes were not described, but cognitive and behavioral modifications (including irritability, delusions, and hallucinations) in isolated ALS patients were already reported in the late 1800's. Additionally, histopathological studies of post-mortem tissues from few ALS cases with dementia provided evidence of anatomical alterations in the cerebral cortex that could explain the cognitive changes observed clinically [14].

However, for a long time, clinicians continued to consider these observations as isolated and atypical cases, convinced that ALS patients remained cognitively normal until the terminal stage of the illness. Only in recent years it has been recognized that ALS is associated to a wide spectrum of behavioral, cognitive or language dysfunctions, which can precede, occur simultaneously or follow the motor symptoms.

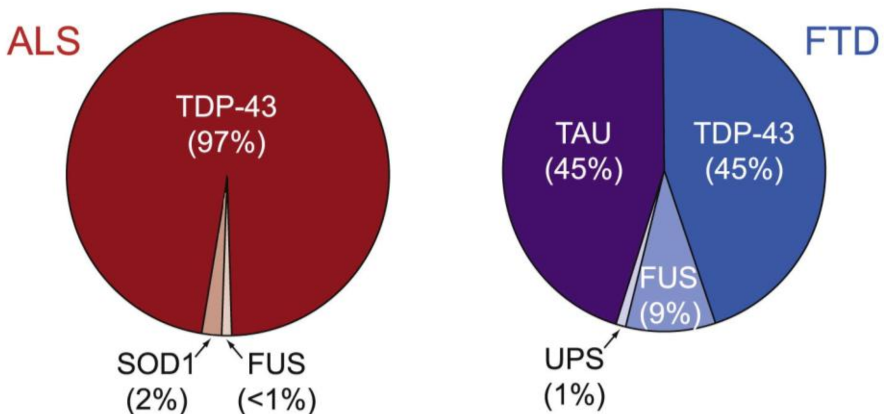
A mild cognitive impairment, affecting executive functions, cognitive flexibility and verbal fluency, is present in 36%-51% of patients with ALS [15] but only a small group of cases (10-15%) shows a more severe dementia, which fulfills the clinical criteria for fronto-temporal dementia (FTD) [16].

Second to Alzheimer's disease (AD) in prevalence, FTD is characterized by the progressive neuronal loss in the frontal and/or temporal lobes leading to a gradual impairment of behavior, language, and personality, with a relative preservation of memory, praxis, and visuospatial functions. The most common clinical phenotype of FTD associated with ALS is the behavioral variant (bvFTD) characterized by early and progressive changes in personality, with disinhibited, apathetic, or manifest stereotypical behaviors [15]. Individuals with ALS-FTD show a tipically onset in their 50's and have a worse survival compared to patients with ALS or FTD alone [17]. In the last few years, this clinical overlap between ALS and FTD has been further supported by neuropathological and genetic findings that have been fundamental to recognize that ALS and FTD represent a disease continuum. In fact, ubiquitinated neuronal inclusions containing TAR DNA binding protein 43 (TDP-43) are the pathological hallmark of ALS (Figure 2) but occur also in a sub-population of patients with frontotemporal lobar degeneration with ubiquitin inclusions (FTLD-U) [18]. Similarly, a FUS pathology has been observed in a rare subgroup of FTLD patients and in LMNs of ALS cases carrying mutation in FUS gene [14]. Moreover, a hexanucleotide repeat expansion in the non-coding region of *C9orf72* gene

has been identified recently as the most common cause of both ALS and FTD [20,21].

In other cases atypical clinical features, such as extrapyramidal signs, objective sensory loss, autonomic dysfunction, cerebellar degeneration, or ocular motility disturbance are reported in association with ALS phenotype. These forms occur rarely and are defined as ALS-plus syndrome [22].

Figure 2. Pathological Overlap of ALS and FTD. Pathological protein inclusions in ALS and FTD, according to the major protein misaccumulated. Inclusions of TDP-43 and FUS/TLS in ALS and FTD reflect the pathological overlap of ALS and FTD. Attached from Ling et al. [17].



1.1.2 Clinical diagnosis

The diagnosis of ALS is based on clinical features, electrodiagnostic testing, and exclusion of other pathological conditions with related symptoms. The El Escorial criteria (EEC), first defined by the World Federation of Neurology in 1993 and then revised in 1998, are the most-commonly used guidelines for the diagnosis of ALS world-wide [23,24].

The diagnosis of ALS requires the presence of UMN and LMN signs in different anatomical segments (bulbar, cervical, thoraco-abdominal and lumbosacral) at neurological examination [8].

Additionally, appropriate neurophysiological and neuroradiological investigations play an important role in supporting the clinical diagnosis. Electrophysiological studies include: a) electromyography (EMG) to detect signs of muscle denervation, and provide evidence of LMNs dysfunction; b) nerve conduction studies (NCS) aimed to exclude other mimic conditions, such as peripheral neuropathy or myopathy [8]; c) transcranial magnetic stimulation producing motor evoked potentials (TMS-MEPs) to detect signs of UMN dysfunction. Neuroradiological investigations of brain and spinal cord are executed to exclude other conditions, like structural lesions, that might cause the same clinical signs, while in some instances muscle biopsy can be performed to reveal denervation atrophy and to exclude the suspect of a myopathy. Genetic testing for mutations in ALS causative genes can be additionally requested in familial cases to confirm a suspected genetic condition.

The spreading of symptoms or signs within a region or to other regions, attested thorough periodic medical examinations, is necessary to demonstrate the disease progression, compatible with a neurodegenerative process.

The revised EEC classify ALS in different categories, according to the number of body regions clinically affected [24]:

- Clinically definite ALS : both LMN and UMN signs in at least 3 body regions
- Clinically probable ALS: both LMN and UMN signs in at least 2 body regions
- Clinically probable – laboratory supported ALS: LMN and UMN signs in 1 body region only, associated to electromyographic signs of active and chronic degeneration in at least 2 extremities
- Possible ALS: signs of LMN and UMN in 1 body region only or UMN signs in 2 or more body regions

To assess the cognitive status of ALS patients, clinicians apply particular neuropsychological tests to investigate a range of different cognitive disturbances, minimizing the limitations imposed by dysarthria, hand weakness and other motor disabilities in ALS. In the last years, a specifically implemented cognitive test, called Edinburgh Cognitive and Behavioural ALS Screen (ECAS), has been developed to determine the presence, severity and type of cognitive and/or behavioural changes in ALS patients [25]. This test results particularly sensitive as investigates different domains: executive functions, fluency, language (maming, comprehension, spelling), memory and visuospatial functions.

1.1.3 Etiology of ALS

ALS occurs prevalently as a sporadic form (SALS), a complex condition determined by the interplay of genetic and environmental risk factors. The role of a significant genetic component is supported by twin pairs studies that estimated the heritability of sporadic ALS at 0.61 [26]. In addition, a small but definite increased risk for ALS was found among first-degree relatives of patients [27]. Age and gender are recognized as non genetic ALS risk factors: older age is positively associated with the risk of developing ALS, and men are slightly more likely to develop ALS than women. Environmental factors that have been associated with ALS risk include cigarette smoking, exposure to heavy metals and pesticides, intensive physical activity and head injuries [28].

Only a small proportion of cases (5-10%) reports a positive family history and is classified as familial form (FALS) [29]. FALS cases are clinically indistinguishable from SALS, but generally they are characterized by an equal male:female sex ratio, an earlier age at onset (47.5 years in FALS vs 56 years in SALS), and a longer disease duration [30].

A systematic review of published epidemiological studies show a geographical difference in incidence of FALS, with a higher value in Finland and lower rates in populations of Southern Italy and of the Balkan Peninsula [29]. Generally, an ALS case is considered familial if at least another family member (within four-degree relative) is affected by the same condition. Considering the absence of a definitive consensus on FALS definition, recently Byrne et al. [31] proposed novel consensus criteria that classify FALS forms in three categories according to family history: a) “definite FALS” for pedigrees with three or more affected members; b) “probable FALS” if another first- or second- degree relative suffers of the same condition; c) “possible FALS” if the other affected member is a distant relative.

Families with ALS frequently show an autosomal dominant transmission pattern, although autosomal recessive and X-linked pedigrees are also described [32]. To date, more than 20 genes explaining the 60% of familial forms have been identified, enlighting the great genetic heterogeneity underlying the disease [33].

1.1.4 Role of enviromental risk factors

Many epidemiological studied were conducted to assess the potential involvement of enviromental factors in the development of sporadic forms and a growing list of potential risk factors was proposed.

A strong role of environmental factors in ALS disease pathogenesis was first supposed by studying Western Pacific endemic foci, where in 1960s local populations showed an extremely high incidence rate of ALS and ALS-related syndromes [5]. The most reliable hypothesis suggests that ALS was induced by a particular chronic dietary consumption of cyanobacterial toxin β-methylamino-L-alanine (BMAA) contained in cycad plants and biomagnified in flying foxes [34]. BMAA was, in fact, detected in brain tissue of ALS-affected individiduals of Chamorro population in Guam, and this compound has proven to induce neuronal cell death in different animal models, including primates [35]. Moreover, in these geographical areas with an unusually high incidence of ALS, analysis of soil and drinking water revealed high levels of mineral content, prevalently aluminum and manganese [36]. The role of such modifiable environmental risk factors is confirmed by a drastic decrease of ALS incidence in Guam over the past 30 years (from 70/100,000 in 1960s to 7/100,000 in 1990s), in association with radical changes in socio-economic conditions, dietary habits and life-style [36]. Interestingly, ALS clusters have been identified also in other geographical regions which are subject to frequent "blooms" of cyanobacteria, including the area near Lake Mascoma in New Hampshire

and the Hérault district surrounding Thau lagoon in Southern France [37,38].

Exposure to chemical neurotoxins (organophosphates) has also been hypothesized to explain the increased risk for ALS reported in certain occupational categories, like agricultural workers and veterans of the 1990-1991 Gulf War, for whom a higher exposure to these compounds is expected [28]. Organophosphates, commonly found in a wide variety of insecticide preparations for agricultural use and present also in nerve agents (e.g. sarin and cyclosarin), act by inhibiting the action of acetylcholinesterase (AChE) in nerve cells.

Together with agricultural pesticides, heavy metals, such as mercury and lead, are the most likely toxicants routinely encountered by humans, being present in air, water, soil and multiple dietary sources. Therefore, the potential pathogenic role of heavy metals in triggering neurodegeneration has received considerable attention. Different approaches have been taken to study metal toxicity in ALS, including the measurement of metal levels in patients' body tissues or fluids and the evaluation of ALS-related clinical symptoms in individuals subjected to heavy metal exposure. Individuals exposed to high doses of mercury and lead display numerous neurological abnormalities, including motor signs resembling ALS, such as extremity weakness, spasticity, hyperreflexia, fasciculation and ataxia [39]. Also animal studies seem to suggest a role of mercury in disease development, as indicated by the observation that a single dose of mercuric chloride ($HgCl_2$) in mice can cause deposition of mercury in upper and lower motor neurons [40]. Moreover, mouse models of ALS (transgenic mice SOD1 G93A) treated with methylmercury show an earlier onset of hind limb weakness [40]. However, a cause and effect relationship between exposure to mercury and ALS has never been specifically demonstrated.

Additional studies focused on different metals, including selenium, copper and zinc, aluminum and cadmium, but they failed to produce reliable and robust results [41].

Other associations have been postulated with intense physical activity, head injury, viral infections, cigarette smoking, electromagnetic fields (EMF) and electrical shocks [28]. The involvement of physical activity as a risk factor has been first hypothesized after that the legendary baseball player Lou Gehrig died from ALS in 1930's and has been recently reinforced by the report of a higher than expected incidence of the disease in professional soccer and football players [42]. However, evidence from epidemiological research remains conflicting and inconclusive. Persistent viral infections have been suggested as a predisposing factor for the development of ALS, after finding enterovirus sequences in a consistent group of spinal cord samples from ALS patients. Some retroviral infections (HIV and HTLV-1) have also been described in association with ALS-like syndromes [43].

Although there is a compelling evidence of the role of environment in ALS susceptibility, the lack of a robust evidence underline the importance of novel studies on larger cohort of individuals to elucidate the direct involvement of exogenous factors in ALS etiology. More recently, the involvement of epigenetic mechanisms (DNA methylation and histone acetylation) has been investigated in an increasing number of neurodegenerative diseases. It's believed that epigenetic changes on the genome may be a mechanism through which environment can contribute to the pathophysiology of age-related neurodegenerative disorders.

An increase in global methylation was observed in postmortem SALS spinal cord, probably reflecting the significant abnormalities of methylation regulators such as the DNMTs (Dnmt1, Dnmt3a) detected in motor neurons [44].

1.2 Genetics of ALS

The genetics of ALS is very complex, characterized by the involvement of major causative genes which underly familial cases and a small fraction of SALS cases, and by susceptibility factors contributing to the etiology of the sporadic forms in association with other low penetrant genetic variants and/or with environmental risk factors [28]. Over the past decade the research in ALS genetics has been extremely productive: it provided fundamental insights into the pathogenesis of the disease unraveling the involvement of multiple and different biological process mediating neurodegeneration and facilitating disease modeling useful for therapeutic testing.

1.2.1 Causative genes

To date, genetic studies have identified more than 20 major genes, whose mutations cause familial forms of ALS (Table 1). These causative genes have a key role in the pathogenesis of the disease, being able to induce ALS with a monogenic Mendelian pattern of inheritance. Both autosomal dominant (*SOD1, c9orf72, FUS/TLS, TARDBP, VAPB, OPTN, VCP, DAO, SETX, hnRNPA2B1/A1*), autosomal recessive (*ALS2, SPG11, OPTN*) and X-linked (*UBQLN2)* forms of ALS have been reported [45]. Thanks to the advent of next-generation sequencing which led to the identification of novel disease causative genes in the last years, now the genetic cause is known for about 60% of FALS. Notably, mutations in major genes associated to FALS have been also identified in patients with apparently sporadic ALS. Some of these patients may be FALS misdiagnosed as SALS for the lack of complete information about other family members. Other mechanisms explaining this finding are incomplete penetrance of the mutations or the occurrence of de novo mutations.

Although the great genetic heterogeneity reported in ALS, the major genetic contributors to FALS are *SOD1, c9orf72, TARDBP* and *FUS,* which globally account for almost 50% of familial and 10% of sporadic forms[45]. For these genes, that have been extensively characterized through large-scale epidemiological studies and supported by functional studies, an accurate understanding in terms of mutational frequency, penetrance and genotype-phenotype correlation is available.

SOD1

SOD1 was the first gene to be identified in 1993, based on linkage analysis in autosomal dominant FALS pedigrees [46]. The gene encode for the Cu/Zn superoxide dismutase, a cytoplasmic enzyme responsible for the catabolism of superoxide radicals to hydrogen peroxide and molecular oxygen. More than 140 different *SOD1* mutations, collectively accounting for ~20% of all FALS patients and in ~3% of SALS cases, have been identified in all the five exons of the gene. All *SOD1* mutations are inherited as dominant traits, with the exception of the D90A variant, that is observed both in recessive pedigrees in Scandinavia, and in dominant pedigrees in the rest of the world [47]. *SOD1* mutations are characterized by a considerable interfamilial and intrafamilial phenotypic variability with regards to the age at onset, site of onset, and disease duration. The A4V mutation, the most frequent in North America, is consistently associated with a high penetrance, younger age at onset, prevalence of lower motor neuron signs, and a short survival with death occurring usually within 12 months. Conversely, other mutations, such as G41D, H46R and G93D, may display a very mild phenotype, with carriers often surviving more than 20 years after disease onset [47].

The identification of *SOD1* was a major breakout in ALS because in the following years the development and the extensive study of different mutant

SOD1 transgenic mice were fundamental to investigate the pathogenic mechanisms and pathways involved in motor neuron (MN) degeneration.

C9ORF72

A hexanucleotide intronic repeat expansion within *C9orf72* gene was identified recently as the most common genetic cause of familial and sporadic ALS and FTD [19-20]. The repeat (GGGGCC) is highly polymorphic in the normal population (2–23 units), but is massively expanded in patients (up to 4000 units). Mutational frequency varies between different populations and countries, ranging from 23%–47% in FALS and 4%–8% in SALS in Western populations [46]. Within Europe, the highest mutation frequency is observed in Finland (46.4% of FALS and 21.1% of SALS), probably a condition determined by the combination of two genetic phenomena (the founder effect and genetic drift) [48]. The contribution of *C9orf72* gene to ALS results significantly lower in non European populations: the repeat expansion was found to account for 2.8% of familial ALS and 0.4% of sporadic ALS in Japanese population [49] and similar low mutational frequencies were observed also in Iran [50]. ALS patients with *C9orf72* repeat expansions are characterized by a preferential bulbar onset, a shorter survival time, concurrence of cognitive impairment and involvement of extra-motor neuronal structures. The finding that individuals with concurrent ALS and FTLD or with a family history of dementia or motor neuron disease have a higher risk of harboring *C9orf72* repeat expansions (33-86%) further support that recognition the two diseases belong to the same pathogenic continuum [51].

The exact cellular functions of C9ORF72, expressed in many tissues, including cerebellum, cortex and spinal cord, remains undefined but a recent study demonstrated that C9ORF72 regulates intracellular trafficking processes in the endosomal and autophagy-lysosomal compartments [52].

TARDBP

The *TARDBP* gene was identified as ALS causative gene in 2008 [53], after the discovery that the gene product, 43-kDa TAR-DNA binding protein (TDP-43) , was a major constituent of ubiquitinated cytoplasmic inclusions in ALS affected motor neurons [21]. TDP-43 is a DNA/RNA binding protein belonging involved in different processes of RNA metabolism, including gene transcription, splicing regulation, and transport and stabilization of mRNA molecules [54]. To date, a total of 47 missense variants and one deletion/insertion have been identified in ALS cases: the current prevalence of these mutations is ~5% of FALS and ~1.5% of SALS cases, with notable differences between different geographical regions [55]. The penetrance of *TARDBP* mutations is believed to be incomplete, with several carriers not developing ALS even in their advanced age. Rarely, patients carrying *TARDBP* mutations can present signs of cognitive impairment, FTD or parkinsonism. With a single exception (the p.D169G mutation), the rest of the pathogenic mutations identified so far are located in the C-terminal glycine-rich region encoded by exon 6 [55]. The p.A382T variant, initially identified in two French cases, is the most commonly observed in patients. This variant, which is also associated to parkinsonisms, displays a high mutational frequency in Sardinia (20-30%) possibly due to a founder effect [56].

FUS

Mutations in FUS were identified in 2009 by two different research groups in ALS families with a previously identified linkage on a locus in chromosome16 (ALS6) [57-59]. The *FUS* gene, comprising 15 exons, encodes for a DNA/RNA binding protein normally localized in the cell nucleus, which plays a similar role to TDP-43 in transcriptional regulation, RNA splicing, nucleo-cytoplasmic RNA shuttling, and maturation of

mRNAs [54]. So far, 58 mutations have been identified with a mutational frequency of approximaly 4-5% in FALS and 1% in SALS [55]. The vast majority of FUS mutations are missense substitutions, but also splicing, frameshit and non-sense mutations have been reported. Similarly to TARDBP, most of the ALS-linked variants of FUS are clustered in the C-terminal region, which contains the nuclear localization signal (NLS) of the protein [54]. Further, FUS mutations have been often associated with a juvenile onset highly aggressive type of ALS.

Mutations in other genes associated with FALS have been reported in isolated families and collectively account for a small proportions of cases (less than~5%) [45]. The genetic cause is still unknown for the remaing 40% of FALS cases.

Table 1. Causative ALS genes

Causative genes	Inheritance	Estimated % of FALS
Enzyme		
Superoxide dismutase 1 (*SOD1*)	Dominant	20%
RNA metabolism		
TAR DNA-binding protein 43 (*TARDBP*)	Dominant	1-5%
Fused in Sarcoma *(FUS)*	Dominant	1-5%
TATA-binding protein associated factor 15 (*TAF15*)	Unknown	Unknown
Ewing Sarcoma breakpoint region 1 (*EWSR1*)	Unknown	Unknown
Angiogenin (*ANG*)	Dominant	<1%
Senataxin (*SETX*)	Dominant	Unknown
Matrin 3 *(MATR3)*	Dominant	Unknown
Repeat Expansions		
C9ORF72	Dominant	22-65%
Ataxin 2 (*ATXN2*)	Dominant	<1%
Proteostatic proteins		
Ubiquilin 2 (*UBQLN2*)	Dominant	<1%
Optineurin (*OPTN*)	Dominant	<1%
Sequestosome (*SQSTM1*)	Dominant	Unknown
Valosin-containing protein (*VCP*)	Dominant	<1%
CHMP2B	Dominant	Unknown
Phosphatidylinositol 3,5-biphosphate 5-phosphatase (*FIG4*)	Dominant	Unknown
Excitotoxicity		
D-amino-acid oxidase (*DAO*)	Dominant	<1%
Cytoskeleton/cellular tran sport deficits		
Vescicle-associated membrane protein B *(VAPB)*	Dominant	<1%
Peripherin *(PRPH)*	Sporadic	Unknown
Dynactin 1 (*DCTN1*)	Dominant	Unknown
Neurofilament heavy chain (*NFH*)	Dominant	Unknown
Profilin 1 (*PFN1*)	Dominant	Unknown
Uncertain		
Spatacsin	Recessive	Unknown
Alsin (*ALS2*)	Recessive	<1%

1.2.2 Genetic risk factors

The great majority of SALS cases are thought to result from the complex interplay between environmental factors and “susceptibility genes”, each one of them increasing the individual’s predisposition to ALS with an additive effect. Many predisposing genes have been identified so far by using candidate association case-controls studies but often the associations were not replicated in other studies.

Among the genes implicated in the etiopathogenesis of sporadic ALS there are *ANG* and *VEGF,* which encode for proteins involved in the formation of new blood vessels (a process known as angiogenesis) but increasingly recognized to act as neurotrophic molecules regulating neurogenesis, neuronal survival and migration [60]. Several lines of evidence, functional and genetic, have associated these genes to ALS. Knockout mice for the hypoxia-response element (HRE) of the vascular endothelial growth factor (*VEGF*) promoter develop an adult-onset progressive motor neuron degeneration resembling ALS. Moreover the intracerebroventricular delivery of recombinant VEGF in SOD1G93A rat model delays onset of paralysis by 17 days, improves motor performance and prolongs survival by 22 days [61]. In humans the single nucleotide polymorphism (SNPs) -2578C/A in the upstream promoter region of *VEGF,* resulting in a lower level of gene expression, has been confirmed by a recent metanalysis to be correlated with an increased susceptibility to ALS in males [62]. Also for the other angiogenic gene, *ANG*, a significant higher frequency of variants has been demonstrated in ALS patients compared to control subjects ($p = 9.3 \times 10^{-6}$) [63].

Many other studies focused on environmental response genes, like *PON* (paraoxonase) and *HFE*, with the aim to find genetic support to the epidemiological evidences of the role of toxin exposure in ALS pathogenesis.

The paraoxonase gene family is a cluster constituted of three 3 genes (*PON1, PON2, PON3*) located in a region spanning ~140 kb on chromosome 7q21.3-q22.1. Their involvement in ALS pathogenesis has been investigated because all three PON proteins share lipid antioxidant properties and moreover, *PON1* is implicated in detoxifying neurotoxic organophosphate compounds [64]. A 5-SNPs haplotype in the *PON1* promoter region was found to be significantly associated with risk of sporadic ALS (p=2.42E-04) and 7 missense mutations in the coding region of *PON1*, predicted to alter protein function, have been identified in FALS and SALS cases but not in controls [64,65].

HFE gene encodes for a membrane protein regulating the iron homeostasis, which play an important role in many biological processes. Given that both iron overload and iron depletion cause neuronal dysfunction, variations in HFE gene have been recently investigated as potential genetic risk factors for ALS and for other neurodegenerative disorders such as Parkinson's disease (PD) [66]. The variant H63D was observed at higher frequency in SALS cases compared to controls in different genetic reports and a recent meta-analysis considering 66,000 cases confirmed that homozygosity for this variant is associated with a 4-fold risk of developing ALS [66].

More recently, intermediate CAG repeats (30-33) in the ATXN2 gene, cause of spinocerebellar ataxia type 2 (SCA2), have been associated with a significant increased risk of ALS (odds ratio (OR) = 4.44) in Caucasian ALS patients [67].

In the last years, genome-wide association studies (GWAs) were applied to identify ALS susceptibility genes in a unbiased manner, but the involvement of multiple loci with an additive effects (polygenic threshold model) complicates the analysis of genetic components contributing to sporadic forms.

1.2.3. Methodologies used to identify ALS genes

1.2.3.1 Linkage analysis

Many ALS pedigrees show a Mendelian pattern of inheritance (mostly autosomal dominant) and therefore can be studied by linkage analysis to map and identify disease-causing genes. Linkage analysis is based on the identification of genetic markers that are co-inherited with the disease and is a powerful method to map disease loci. In the past, researchers used as informative DNA markers Short Tandem Repeat (STRs) or Variable Number Tandem Repeats (VNTRs). Today the analysis is based on SNPs that, although less polymorphic, are more abundant and can be simultaneously tested in automated high-throughput genotyping arrays. A main advantage of linkage analysis is that it is unbiased and genome-wide, and does not require a prior knowledge of the pathogenic mechanisms underlying the disease. Family-based studies, supported by linkage analysis, have been the starting point for the identification of different ALS genes. In 1989 the first locus (ALS1) associated with dominant ALS was identified to be on chromosome 21q22.1 by linkage analysis in 6 multigenerational ALS families [68]. The maximum lod score was obtained for the marker D21S223 localized within exon 2 of *SOD1*. Subsequently, 11 different *SOD1* missense mutations were identified in affected members of 13 familial ALS pedigrees by single-strand conformational polymorphism analysis and direct sequencing of exons [44]. Following this first success, classical linkage studies have led to the discovery of 10 new ALS loci, including loci responsible for recessive forms of ALS, such as ALS2 (chromosome 2q33), caused by mutations in the Alsin gene, and ALS5 (chromosome 15q15-21) caused by mutations in Spatacsin gene [47].

1.2.3.2 Candidate gene approach

The discovery of neuropathological features and pathomechanisms of ALS pointed researchers towards investigating suitable "candidate genes" potentially playing an important role in disease pathogenesis. This approach was successful in identifying *TARDBP*, one of the most common causative genes in ALS. Following the discovery of TDP-43 as a major component of ubiquitinated inclusions in affected motor neurons [18], three different groups screened the entire coding region of *TARDBP* gene in a large cohort of ALS patients and identified different missense mutations [53,55]. Following these studies, an increasing number of reports on ALS patients of different geographic origin were published and effectively *TARDBP* mutations were found to be a major cause of FALS and also of a small fractions of sporadic cases [55].

Similarly, the genes *NEFH*, encoding for the neurofilament heavy subunit, and *PRPH*, encoding for peripherin, were also analyzed as candidate genes because of an abnormal accumulation of intermediate filaments in the perikarya and proximal axons of motor neurons in ALS patients [69]. Six deletions within the C-terminal lysine-serine-proline (KSP) phosphorylation domain of the *NEFH* gene were found in several SALS patients and in a pedigree with autosomal dominant FALS, while an homozygous missense mutation and a heterozygous frameshift deletion of the *PRPH* gene were observed in ALS cases [47].

Finally, other studies selected candidate genes for genetic screening because of structural and functional similarity to known ALS-genes: one example is represented by TAF15 and EWS which together with FUS belong to the FET family of RNA-binding proteins and are implicated in regulation of gene expression, maintenance of genomic integrity and mRNA/microRNA processing [70].

1.2.3.3 Genome-wide association studies

One of the most used approach to identify susceptibility genes in sporadic ALS is represented by genome-wide association (GWA) studies. Unlike candidate-gene approaches, genome-wide association (GWA) study is a non-hypothesis-driven method which analyzes the entire genome (genic and intergenic regions) to find loci associated to the disease. Basically, GWA studies are case-control studies that, using data obtained by high throughput genotyping platforms, compare allelic frequencies of many common SNPs (300.000-600.000) across the human genome between affected individuals and healthy controls. The result of a GWA study is the identification of SNPs associated to the disease, either because they have a direct causative role in the pathogenesis of the condition (coding variants affecting protein function or variants affecting transcription and translation) or because they are in linkage disequilibrium with the actual causal variant. Over the last years, many GWA studies have been performed in large SALS cohorts, identifying several loci associated with disease risk, such as *DPP6* (dipeptidyl peptidase 6), *ITPR2* (inositol 1,4,5-triphosphate receptortype 2), *ELP3* (the elongator protein 3), *SUNC1*, chromosome 9p21, *UNC13a* [45]. Unfortunately, with the exception of the signal on chromosome 9p21, the other genetic associations failed to be robustly replicated in independent studies [71]. The 9p21 locus associated to sporadic ALS overlapped with a 3·6 Mb region previously linked to ALS-FTD in seven families [72]: only after an extensive search, the genetic cause of chromosome 9-linked FTLD-ALS was identified in the non-coding hexanucleotide repeat expansion within the gene C9ORF72 [20,21]. This finding has underlined the critical importance of very large sample sizes in order to have sufficient power to capture allelic association with small effects and low MAF. Recently, by performing a meta-analytical GWAS study on a total of 13,225 individuals (6,100 cases and 7,125 controls) for

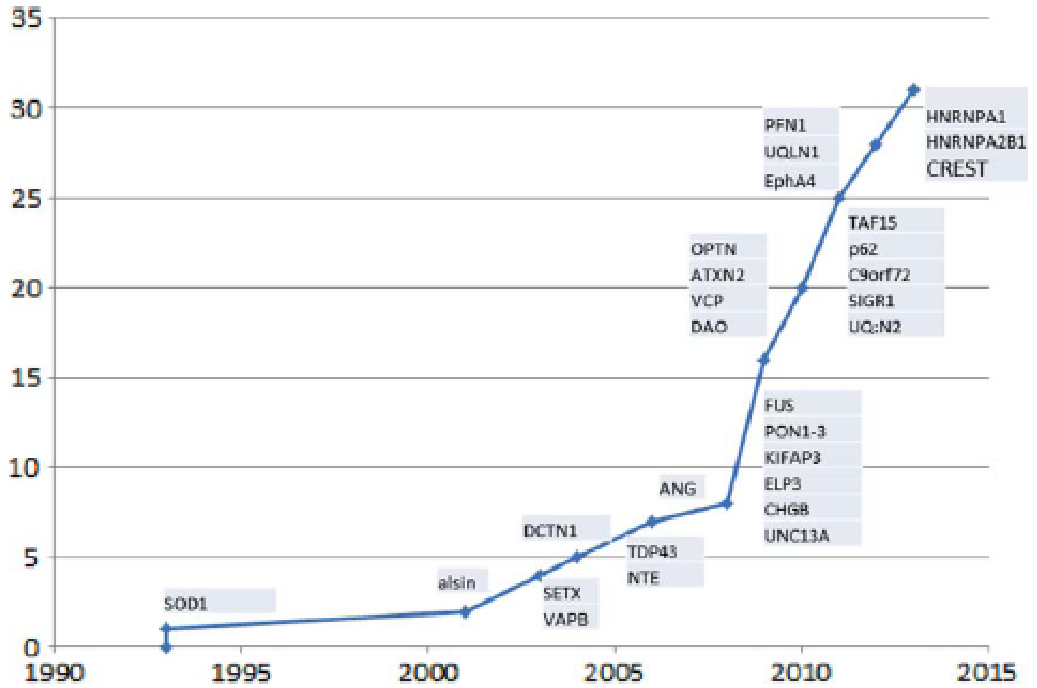
almost 7 million SNPs, our research group confirmed the previously reported association at 9p21 with the expanded hexanucleotide repeat in the C9orf72 gene (rs3849943 with *P* = 7.69 x 10-9; OR 1.16) and identified a novel locus for sporadic ALS risk at 17q11.2 (rs34517613 *P* = 1.11 x 10-8; OR 0.82) that was validated using genotype data from a replication cohort [73]. The lead SNP at this locus was in linkage with three SNPs in the SARM1 gene, which seem to promote an axonal self-destruction pathway known as Wallerian degeneration.

Additionally, GWAs have been also used to identify modifier genes, that not increase the risk of developing ALS but modulate penetrance, age of onset, rate of progression, survival. One example of disease modifier is *KIFAP3*, a gene encoding for a kinesin associated protein involved in cellular anterograde transport and chromosomal cytokinesis. A genome-wide analysis in 1,821 sporadic ALS cases and 2,258 controls revealed that the SNP rs1541160 within the *KIFAP3* gene is associated with reduced KIFAP3 expression and longer survival. The homozygous genotype C/C of the SNP rs1541160 increase the mean survival of ≈14 months, a variation of clinical importance for a rapidly progressive disorder like ALS characterized by a survival of only 3–5 years [74].

1.2.3.4 Whole-exome sequencing

Until recently, it was not feasible to perform genome-wide sequencing on large cohorts. However, the recent advances in automated short-read DNA sequencing technologies have rendered the whole genome sequencing and in particular the target sequencing of the exome a powerful and cost-effective tool for genetic studies. The exome (the protein-coding regions of the genome) constitutes only approximately 1% of the human genome, but harbours about 85% of the mutations responsible for Mendelian disorders. Ever since the first application in Miller syndrome, whole-exome sequencing (WES) has proven to be highly successful in discovering the genetic bases of several rare Mendelian disorders in which conventional approaches have failed [75]. To date WES has proven to be useful also in ALS with the identification of different causative genes (Figure 3). This approach was successful also in small pedigrees, where a limited number of patients was sequenced. For example, the *VCP* gene, previously associated to Inclusion Body Myopathy, Paget disease, and Frontotemporal Dementia (IBMPFD), was identified as an ALS causative gene by performing exome-sequencing on only two affected members of a four-generation Italian ALS family [76]. By applying a similar approach, a mutation in the *hnRNPA2B1* (heterogeneous nuclear ribonucleoprotein A2B1) gene was identified in one family with IBMPFD/ALS, and two mutations in *hnRNPA1* (heterogeneous nuclear ribonucleoprotein A1) gene were found in two families with pure ALS [77] More recently, the exome-sequencing approach has been applied also to sporadic ALS in a “trios” study design (one affected proband and the two unaffected parents) to identify potential *de novo* mutations. This study analyzed 47 ALS trios and reported a significant enrichment of *de novo* mutations in genes encoding for proteins belonging to the chromatin regulatory pathway [78].

Figure 3. Timeline of gene discoveries in familial and sporadic ALS. The exponential growth in genetics related to ALS after the application of exome-sequencing approach. Attached from Renton et al. [79].



1.2.4 Problems and limitations in the study of ALS genetics

The information provided by genetic studies toward the understanding of ALS pathogenesis has been invaluable. It has led over the years to the identification of multiple and different biological process involved in motor neuron degeneration, has provided potential therapeutic targets, and facilitated the engineering of ALS animal models, essential for validating drugs efficacy [1]. Although the great advances achieved over the last decades, only 50-60% of the genetic variability in FALS has been explained so far, while the genetic component of sporadic forms is even less understood. This is explained in part by the complex genetic heterogeneity underlying ALS and by the limited success achieved by traditional genetic approaches, such as the linkage analysis and the candidate gene approach. Linkage analysis requires the study of large pedigrees composed of many affected individuals in different generations. This task is considerably difficult in ALS, that occurs in adult life and has a rapid disease course. Many FALS pedigrees collected for research purpose are not large enough to effectively map the disease gene with the linkage approach. Moreover, similarly to other complex neurodegenerative diseases, the analysis of FALS pedigrees is complicated by a variable age of onset, incomplete penetrance and phenotypic heterogeneity. Misdiagnosis of ALS in family members and pleiotropy of implicated genes can be other sources of error [45]. The candidate approach, which has the advantage to be feasible in cohorts of unrelated FALS and in SALS, is flawed by a selection bias and the several studies conducted so far have been inconsistent and frequently underpowered. Many GWAs studies, aimed at identifying susceptibility genes in SALS, were not able to produce convincing and replicable results. Moreover, odds ratios associated with putative risk alleles identified so far are generally low, indicating a modest risk effect and the probable involvement of many susceptibility genes that

cumulatively elevate the risk of developing ALS. The common variants interrogated by SNPs genotyping chip in GWAs studies seem typically to explain only a small fraction (12-21%) of the calculated inherited contribution to ALS risk [66]. Therefore, it's been proposed that low-frequency or rare variants, not captured by the SNPs arrays used for GWAs but with presumably higher effect sizes, may account for the "missing heritability" in SALS. New approaches (such as genome-wide and exome sequencing) are thus required to define more comprehensively the inherited basis of the disease both for FALS and SALS. Next-generation sequencing and in particular WES have proved to be a successful and efficient method for the study of genetics in neurodegenerative disorders, including ALS. Using this technology researchers have been able to identify causative genes in familial pedigrees analyzing a limited number of affected members and more recently applying a trio–based approach, investigated the contribution of *de novo* mutations to disease susceptibility.

1.3 Pathogenic mechanisms

While the exact molecular and cellular basis for motor neuron degeneration is not yet fully understood, increasing evidence point to the contribution of multiple aberrant biological processes. The cellular events contributing to the pathobiology of ALS include protein aggregation, dysregulated RNA metabolism, oxidative stress, mitochondrial dysfunction, excitotoxicity, impaired axonal transport and neuroinflammation (Figure 4).

1.3. 1 Protein aggregation

Similarly to other neurodegenerative disease (Alzheimer's, Parkinson's and Huntington's diseases), misfolding and protein aggregation seem to be crucial also in ALS pathogenesis. The presence of ubiquitinated inclusions in LMNs of the spinal cord and brainstem and in corticospinal UMNs is the pathological hallmark of ALS. Using a combined biochemical and immunoistochemical approach, TDP-43 was identified as the main component of ubiquitinated cytoplasmic inclusions in ALS and in a sub-population of Frontotemporal lobar dementia with ubiquitin inclusions (FTLD-U) patients [18]. In aggregates, TDP-43 is hyperphosphorylated and cleaved to generate abnormal C-terminal fragments. Since in ALS affected motor neurons TDP-43 is mislocalized in the cytoplasm forming intracellular aggregates, its altered localization may play a pivotal role in neurodegeneration resulting in the loss of their proper function in the nucleus (loss-of-function effects) and/or in their potential toxicity in the cytoplasm (gain-of-function effects) [54]. Misfolding and aggregation is a propensity demonstrated also for other ALS-linked proteins, including FUS, SOD1, OPTN, UBQLN2, HNRNPA2/B1, VCP and the poly-(glycine-proline) peptides generated by repeat associated non-ATG translation (RAN translation) of the C9ORF72 repeat transcripts [1]. Cytoplasmatic inclusion immunoreactive for FUS have been found in disease affected

tissues from patients harboring FUS mutations [80] but , in contrast to TDP-43, there are no evidence of the presence of hyperphosphorylated or cleaved forms of FUS protein in the aggregates. Interestingly, both TDP-43 and FUS contain 'prion-related' glutamine/asparagine (Q/N) rich domains that could induce protein aggregation [81]: a similar domain is present in two other RNA-binding protein, hnRNPA2/B1 and hnRNPA1, found to be mutated and aggregated in ALS [77]. SOD1 is a primary component of protein aggregates found in SOD1-mutated FALS patients and in mutant SOD1 rodent disease models. According to in silico and experimental studies, ALS SOD1 variants show an increased propensity to form aggregates, due to dimer destabilization and dissociation into monomers which then nucleate the formation of aggregates. Protein instability and aggregation propensity are two factor correlating with the disease severity in SOD1+ FALS patients [82].

1.3.2 RNA metabolism

Abberant RNA has been described in many neurodegenerative diseases and seems to be of major importance also in ALS pathogenesis. TDP-43 and FUS are two RNA-binding proteins involved in multiple steps of RNA processing pathway, from transcription to post-transcriptional regulatory process (splicing, mRNA stabilization and transport). The altered localization of these two proteins forming cytoplasmic inclusions in affected tissue determines a sequestration and loss of their normal nuclear function with possible severe consequences on mRNA metabolism and therefore on the survival of motor neurons.

The role played by altered RNA homeostasis in ALS pathogenesis is further emphasized by the evidence that several mutations observed in patients affect other RNA-processing genes, including *hnRNPA2/B1*, optineurin (*OPTN)*, *senataxin* (*SETX)*, *angiogenin (ANG)*, *elongator protein 3* (*Elp3*) and *survival motor neuron* (*SMN*) [54]. The recent finding that sense and

anti-sense transcripts of the expanded repeat in C9orf72 accumulate in nuclear RNA foci, sequestering several RNA-binding proteins, has further supported the central role of RNA dysregulation in ALS [83].

1.3.3 Mitochondrial dysfunction

Several experimental evidences support the notion that alterations in mitochondrial morphology, bioenergetics and fusion/fission dynamics play an important role in ALS [84]. Morphological alterations, such as fragmentation, and changes in mitochondrial function, like a reduction in the activity of respiratory chain complexes and decline of mitochondrial bioenergetic capacity, have been reported in ALS patients, in cultured cells expressing mutant SOD1 and in mutant SOD1 transgenic mice. Although a cytosolic protein, a fraction of SOD1 (1-2%) is translocated into the mitochondrial intermembrane space (IMS) with the function to protect the cell from additional oxidative stress damage by oxygen radicals. An increased mitochondrial recruitment and accumulation have been observed for mutant SOD1: mitochondrial mutSOD1 may directly damage these organelles by forming toxic aggregates or by engaging abnormal interactions with other mitochondrial proteins, like Bcl-2 [52].

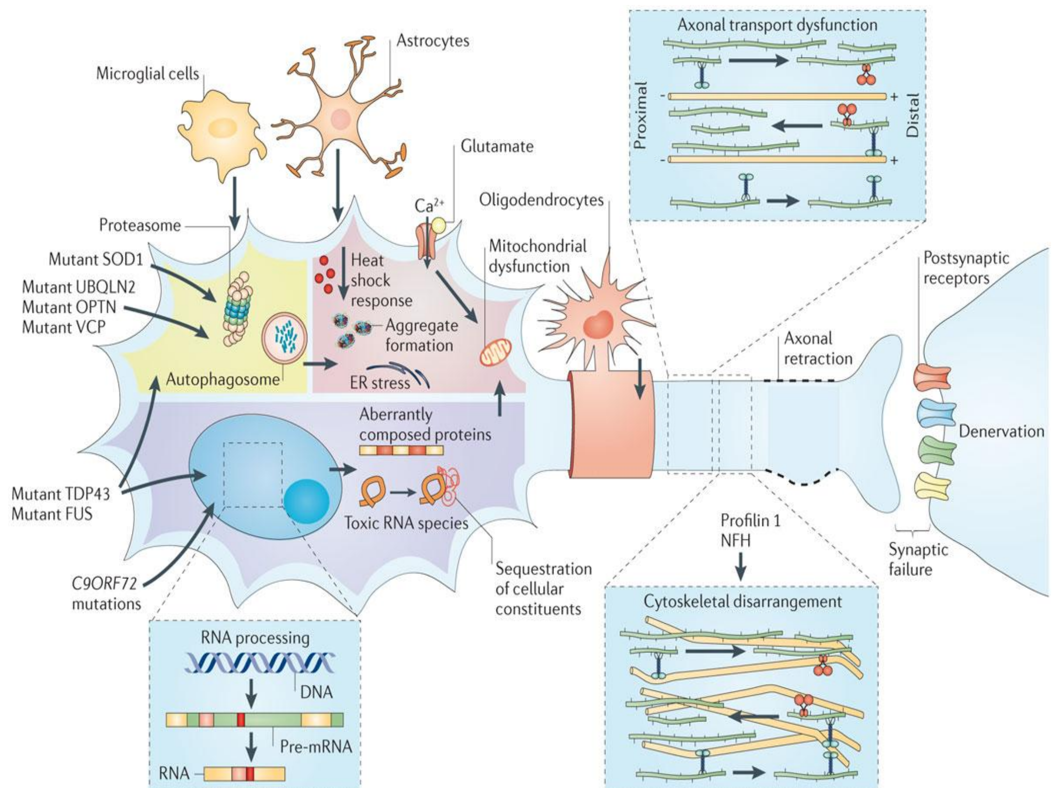
1.3.4 Cytoskeletal abnormalities

Different cytoskeletal abnormalities have been reported in ALS cases, including abnormal accumulation of intermediate filaments (IF) in the perikaryon and axons of motor neurons. Neurofilaments (NFs) are the major type of IFs in adult motor neurons and are formed by three proteins, NF light (NF-L), NF medium (NF-M) and NF heavy (NF-H). Genetic defects in the NF-H gene have been identified in a small number of patients, providing compelling evidence for involvement of cytoskeletal defects in ALS. Significant loss of motor axons (20 to 25%) has been reported in NF-L null mice and in mice knockout for both NF-M and NF-H [85], while the

overexpression of any of the three NF subunits in mice induced the formation of perikaryal NF accumulations reminiscent of those found in human ALS. Another type of IF detected in the abnormal inclusion in ALS motor neurons is peripherin. Notably, the expression of peripherin can be upregulated by imflammatory cytokines, such as IL-6, which was found to be increased in the cerebrospinal fluid of ALS patients.

Other potential mechanisms of toxicity targeting motor neurons defects in axonal transport, glial cell pathology, and glutamate excitotoxicity, although the primary or secondary role of each of these events in triggering MN degeneration still need to be determined [84].

Figure 4. Pathogenic mechanisms in ALS. Attached from Robberecht et al. [84]



2. AIM OF THE PROJECT

Despite numerous advances over the last few years, the genetic basis of ~40% of familial ALS remains to be identified, while the genetic causes of sporadic ALS are even less understood. Continuous efforts are needed to identify the missing heritability and to increase our understanding of the molecular pathogenic mechanisms underlying the disease.

The main aim of this research project was to identify novel genes associated to familial and sporadic forms of ALS by applying complementary approaches based on next-generation sequencing, thus overcoming the limitations of traditional genetic strategies.

In particular we addressed two specific aims:

- Identification of novel causative genes in familial ALS.

  Exome-sequencing was perfomed on familial cases, negative for mutations in known ALS-genes and collected through collaborations with other national and international ALS centers. The cohort included two ALS pedigrees displaying a dominant pattern of inheritance and 363 unrelated index FALS cases (one affected member for family), to reach an adequate number of affected individuals for a statistically powerful study.

- Identification of potential DNMs in sporadic ALS

  A set of 32 ALS trios underwent exome-sequencing, according to the hypothesis that *de novo* mutations may contribute to the genetic basis of sporadic forms.

3. MATERIALS AND METHODS

3.1 Patients and controls

The patient cohort consisted of patients with familial or sporadic ALS, collected from different centres in Italy, Europe e USA. All included patients fulfilled the El Escorial revised criteria [24] and were classified as having FALS according to the criteria of Byrne et al. [31]. Genomic DNA was extracted by peripheral blood after informed consent was obtained and all protocols were approved by the Institutional Review Boards at the institutions involved. FALS cases were prescreened for mutations in the following ALS-causative genes, *SOD1, FUS, TARDBP*, *C9ORF72, VCP* and *ANG* genes, and individuals harbouring causative mutations were excluded from the study. SALS samples were analyzed for *SOD1, FUS, TARDBP, C9ORF72.*

In particular, the cohort collected for exome-sequencing studies included:

- Two large ALS families of Caucasian (Family #1, Figure 5A) and Sephardic Jewish (Family #2, Figure 5B) origin, displaying a dominant pattern of inheritance. DNA was available for 4 affected and 13 unaffected members of Family#1 and for 17 individuals (9 affected and 8 unaffected) of Family#2. In each family, the two affected members with maximum genetic distance were selected for exome-sequencing.

- 363 index unrelated FALS cases collected from several European countries and from USA (Table 2).

- 32 ALS trios, each constituted by the patient and both unaffected parents. The SALS group included 10 females and 22 males, with an average age at onset of 41,0 ± 9,3 years (median 42,0 years), and a mean survival after the onset of first symptoms of 41,9 ± 21,7 months. None of the selected patients had cognitive impairment or dementia. Average age at

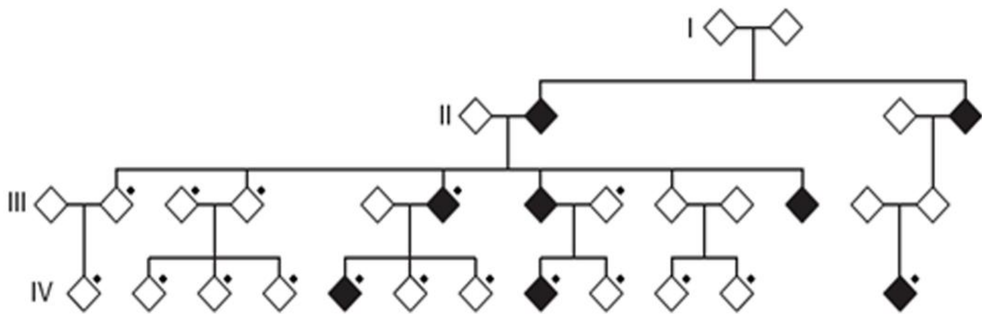
sampling for the unaffected parents was 74,8 years (72,0 for females and 77,4 for males). The clinical characteristics of the enrolled patients are reported in Table 3.

Moreover 31 controls samples (Table 2), ethnically-matched, and 9 FALS with known causative mutations (6 in SOD1 and 3 in VCP) were subject to WES.

A second group of 272 FALS devoid of known mutations underwent to exome-sequencing and was used as replication panel. Additional FALS, SALS samples and FTD patients (fulfilling the Neary criteria for FTD diagnosis) were included in the study for mutational screenings aimed to validate candidate ALS genes.

Figure 5. Pedigrees of Family #1 and Family #2. Affected members are shown as black colored boxes, while the black dots indicate the family members whose DNA was available for the genetic study. Sex of pedigree members is obscured to protect privacy. Both ALS familial pedigrees show a dominant pattern of inheritance (A, B).

A Family#1



B Family#2

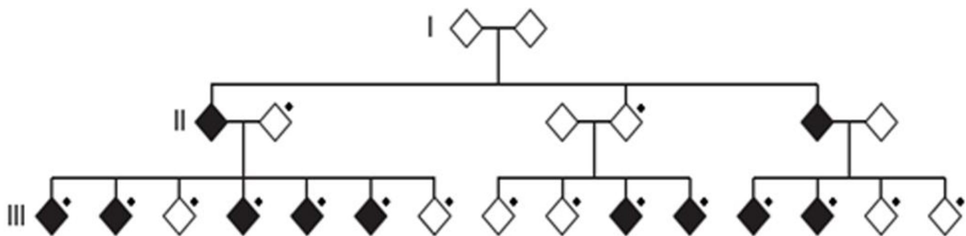


Table 2. Index FALS cases and controls collected for exome-sequencing. All familial cases recruited resulted negative for mutations in known ALS-associated genes.

	FALS	Controls
Ireland	16	6
Italy	97	6
Spain	29	6
United Kingdom	140	7
United States	81	6
Total	363	31

Table 3. Demographic and clinical features of ALS trios.

Patients	N°	Age at onset	Site of onset	
			bulbar	spinal
M	**22**	**41.9**	**1**	**21**
F	**10**	**42.6**	**3**	**7**
Total	**32**	**41.8**	**4**	**28**

3.2 Linkage analysis and exome-sequencing

Linkage analysis

Linkage analysis was perfomed on the available family pedigrees, Family#1 and Family#2, by genotyping family members with Affymetrix 10K SNP arrays and analyzing data with the software application Allegro v.2.031 to generate multi-point LOD scores. Only affected members (and married-in samples) were used for the analysis.

Exome-sequencing

The targeted capture from genomic DNA was performed using the Nimblegen SeqCap EZ Exome Library, that containing 2.1 million long oligonucleotide probes provide the highest density coverage of RefSeq Coding Exons compared to other enrichment platforms (Illumina or Agilent). For FALS index cases and ALS trios samples, unique short sequences (barcodes) were attached to the adaptor region during the exome capture process to allow for multiplex sequencing. Short-read sequencing was performed on a HiSeq2000 instrument (Illumina) with an average coverage of 50X per exome. Exome sequencing reads were aligned to a human reference (hg19) using BWA (Burrows-Wheeler Aligner) and processed using the Genome Analysis ToolKit (GATK). After removal of duplicate reads, index realignment (GATK IndelRealigner), and base quality recalibration (GATK TableRecalibration) was performed. Variant detection and genotyping were performed using the UnifiedGenotyper (UG) tool from GATK. Variants not passing quality control criteria were eliminated (QD<5.0, HRun>3, MQ<40.0, FS>60.0, HaplotypeScore>13.0, MQRankSum< -12.5, ReadPosRankSum< -8.0) and genotypes with low quality (GQ<50) were set to missing.

3.3 Data analysis strategies

Data analysis and variant filtering pipelines were handled differently for the three subsets (FALS pedigrees, index FALS and ALS trios).

FALS pedigrees

To reduce the number of candidate variants from the initial set identified in every individual by WES, different filtering steps were applied, as reported in previous exome sequencing studies [76,77]. Based on the main assumptions that the causal variant alters the protein-coding sequence and is not present in the general control population, the non-coding variants or SNPs identified in dbSNP132, the 1000 Genomes Project (May 2011 release), or the NHLBI ESP Exome Variant Server (5,379 sequenced exomes) were excluded. We then filtered out variants that were not shared by both affected members of each family or variants not in according to the dominant inheritance pattern or excluded by linkage analysis. The steps are shown in Table 5. Candidate variants identified by filtering were validated using bidirectional Sanger sequencing with customized primers in the patients with exome sequence data and tested for segregation in the additional affected members of the Family#1 and Family#2.

FALS unrelated

Samples with low call rate(<0.75), showing relatedness to another sample (IBD>0.2), with discordant phenotype-genotype gender, with excess homozygosity/heterozygosity or outliers from stratification analysis were eliminated resulting in the final set of FALS samples used for the next analysis.

A rare variant analysis was performed by logistic regression of case-control status with respect to the aggregated count of minor alleles in a given gene window. Regression was performed using Firth's penalized likelihood method to avoid errors of model fitting in the event of data separation

(where gene variants occur exclusively in the case or control cohort). For controls, genotype data from 4300 European Americans were obtained from the NHLBI's Exome Variant Server. These genotypes were randomly assigned to 4300 simulated individuals. Variants were filtered to those showing 90% call rate in either ESP data set or FALS. Gene selection was based on the initial set that passed quality control standards of the ESP study (14,931 autosomal genes). These genes were further restricted to those which displayed an 85% call rate over 85% of the coding region in the EVS resulting in the final set of 12,224 genes subject to rare variant analysis.

Variants were annotated as damaging by PolyPhen2 (score> 0.446) or representing stop gain/loss variants. Additional filtering procedures included SIFT ("damaging"), Mutation Taster ("disease causing automatic" or "disease causing"), GERP (GERP++ score>0), and phyloP (score>0). Variants that did not display a 90% call rate in either cases or controls were eliminated from the analysis. Gene windows and predicted variant effects were defined based on protein coding RefSeq transcripts only. Variants were only included for analysis if the associated minor allele frequency did not exceed what would be expected of a typical ALS mutation on the basis of prior publications. Specifically, Mendelian ALS variants identified to date have generally been reported to account for <1% of population based patient cohorts8,9. The lifetime risk of ALS is ~1/40039, meaning that even in the event of penetrance as low as 50%, it would not be expected to observe a given ALS mutation in more than a single control (p<0.05). To allow for the potential of a single control occurrence and up to 3 case occurrences (~1% of patients), a maximum minor allele frequency of 0.0004 (4 occurrences) was imposed. Multiple testing correction was performed through repetition of regression analyses with prior permutation of case-control status. Reported family-wise error rates represent the

proportions of permuted datasets wherein any gene exhibited an association of equal or greater significance. The genomic inflation factor was calculated as previously described. All data processing and rare variant analyses were performed using scripts prepared for bash version 4.2.24, python version 2.6.5 and R version 3.0.1.

Trios samples

On each trios sample, paternity and maternity was verified by checking the percentage of shared heterozygous mutations between child and mother/father. Predicted *de novo* events were identified as sites where both parents were homozygous for the reference sequence and the offspring was heterozygous. To identify rare private variants, the full variant list was compared against dbSNPv135, the NHLBI Exome Sequencing Project and the 1000 Genomes Project.

3.4 Bioinformatic analysis

The effect of missense variants on protein structure and function was predicted by using different bioinformatic tools: SIFT (Sorting Intolerant from Intolerant), an evolutionary conserved based method, and PolyPhen-2, a prediction tool which evaluate also the physic-chemical characteristics of the amino-acid involved in the substitution and other structural parameters. The gene function analysis and the functional classification according to the molecular function was performed by using Panther (Protein ANalysis THrough Evolutionary Relationships). Two different gene network tools, DAPPLE (Disease Association Protein-Protein Link Evaluator) and GeneMania, were used to construct networks using the list of genes of interest and the subset of known ALS-causative genes.

3.5 Sanger sequencing and genotyping

Sanger sequencing was applied to samples to validate the candidate variants indentified by WES and to exclude possible false positive. Further, the mutational screening of candidate genes *(PFN1, TUBA4A)* was performed by direct sequencing on additional ALS cases, FTD patients and controls.

The entire coding regions of *PFN1* and *TUBA4A* and at least 30 bp of the flanking intronic sequences were amplified by polymerase chain reaction (PCR) using AmpliTaq Gold (Applied Biosystems) or AccuPrime GC-rich Polymerase (Invitrogen). Amplification primers are shown in Table 4.

The PCR products were subsequently purified by incubation with Illustra Exonuclease I and Alkaline Phosphatase (GE Healthcare) to remove unincorporated primers and dNTPs, and then sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and run on a capillary sequencer (Applied Biosystems). All identified sequence variations were confirmed by sequencing an independent PCR product.

Control DNA samples were tested by direct sequencing or by SNP genotyping using TaqMan Assay.

Table 4. Primer sequences and PCR conditions for amplification of *PFN1* and *TUBA4A* coding regions.

Gene	Exon	Primer sequence (5' → 3')	Amplicon length (bp)	Annealing temperature (°C)
PFN1	1	Fwd: CAGCTCGAGCCCAGTCC Rev: AAGTCCCTCCCTCAGGGTC	254	60
	2	Fwd: GGAGAACACGGTGGGAATC Rev:CACCCTCAAGATTACCAGAAGG	330	58
	3	Fwd: GGGTGGAATTGTGACACCTG Rev: TTTGTGTGTGTATGGGGAGG	223	58
TUBA4A	1	Fwd: CAGCCTCTAGTGGGCGAGC Rev: CTCGCACCTCCTGGAGAC	262	60
	2/3	Fwd: AGCTGGTAGATGATAGGGTGG Rev: GATGACCTCCTGAAAGGATCTG	845	60
	4	Fwd:CTGACTGCTGATGTATCTTACGC Rev:AGCAGAAGCTCAAGCACTCAGAGG	1175	61

3.6 Cell cultures and transfections

The mouse neuroblastoma cell line N2A was grown in MEM supplemented with 10% fetal bovine serum, 2mM L-glutamine, 100 units/ml penicillin and 100 µg/ml streptomycin at 5% CO_2. HEK293 cells were maintained in DMEM supplemented with 10% fetal bovine serum, 100 units/ml penicillin and 100 µg/ml streptomycin at 5% CO_2.

Primary motor neurons (PMNs) were isolated from E12.5 mouse embryos: spinal cords were dissociated in 0.1% trypsin for 12 minutes at 37ºC, and motor neurons were isolated by density centrifugation on a 6% Optiprep (Sigma-Aldrich) cushion. Primary motor neurons were maintened in Neurobasal medium supplemented with 2% horse serum, 2% B27, 10nM BDNF, CNTF, GDNF).

PFN1 expression vectors with V5 epitope tags were constructed using the backbone pcDNA3.1/nV5-DEST (Invitrogen) by way of Gateway Technology according to manufacturer's protocol. For establishment of mutant Profilin1 plasmids, site-directed mutagenesis was performed on wild-type vector using Quickchange Multi Site-Directed Mutagenesis Kit (Agilent) according to manufacturer's protocol. TUBA4A expression constructs were created by synthesizing the wild-type and mutant coding region with an HA epitope tag at the N-terminus (DNA2.0) and subcloning into the vector pJ603.

For Western Blotting experiments, transfections of N2A and HEK293 cells were performed in 6-well plates with 3 µg of plasmid DNA using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. For immunofluorescence, cells were plated in 12mm round coverslips in 24-well plates and transiently transfected with 0.8 µg of plasmid DNA plasmids using Lipofectamine 2000 (Invitrogen).

PMNs were transfected using paramagnetic nanobeads (NeuroMag, Oz Biosciences) as previously described [86].

3.7 Immunofluorescence

At 48h after transfection, N2A cells were fixed with 4% paraformaldehyde for 15 minutes, permeabilzed with 1% Triton X-100 at room temperature for 10 minutes, and then blocked with blocking buffer (50 mM NH4Cl, 10 mg/ml BSA, 2% natural goat serum, 0.1% Triton X-100 in DPBS) at 37°C for one hour. Cells were then incubated with a fluorescent-conjugated primary antibody (mouse Dylight 549-V5 antibody, rabbit Dylight 488-ubiquitin antibody) at 4°C overnight. After the incubation, cells were washed three times with PBST (0.1% Tween-20), and the coverslips mounted with Vectashield Hard Set Mounting Medium containing DAPI (Vector Laboratories). Similarly, immunofluorescence experiments were also conducted on primary motor neurons transfected by magnofectin.

HEK293 cells were fixed 48-hour after transfection with 4% paraformaldehyde at room temperature for 15 minutes, permeabilized with cold methanol for 10 minutes and 0.3% Triton X-100 at room temperature for 5 minutes followed by blocking with 10% normal goat serum solution in PBS at room temperature for 20 minutes. Incubation with primary antibody (anti-HA, 1:100, Sigma) was performed in blocking solution for 1.5 hour at 37°C and the fluorescent secondary antibody AlexaFluor 488 (1:500, Invitrogen) was used for detection. Slides were mounted with ProLong Gold antifade reagent containing DAPI to visualize the nuclei.

Confocal images were obtained by Leica or Nikon TE-2000E2 microscope.

3.7 Soluble/ insoluble fractionation

NP40 soluble/insoluble fraction

After transfection, N2A cells were directly collected in NP-40 lysis buffer (1% NP-40, 20 mM TrisHCl pH. 7.4, 150 mM NaCl, 5 mM EDTA, 10% glycerol, 1mM DTT, 10 mM sodium fluoride, 1 mM sodium orthovanadate, 5 mM sodium pyrophosphate) additioned with EDTA free protease inhibitors (Roche). After sonication, the lysate was rotated for 30 minutes at 4°C followed by centrifugation at 13,500 rpm for 20 minutes. The supernatant ('soluble fraction') was collected into a new tube, while the remaining pellet was washed once with 1 ml of lysis buffer and then resuspended in 8M urea/3% SDS buffer followed by sonication ('insoluble fraction'). The soluble/insoluble fractions were analzyed by Western Blotting using the mouse monoclonal anti-V5 (1:2000, Invitrogen) antibody to detect PFN1. GAPDH was used for normalization.

Differential solubilization

TUBA4A transfected HEK293 cells were collected at 48h in 300µl PBS buffer with protease inhibitors (Sigma), sonicated and centrifuged at 13,500xg for 10 minutes at 4°C. The supernatant was collected as the PBS-soluble fraction. The remaining pellet was sequentially resuspended in 100µl 1% Triton X-100, 5% SDS, and 8M urea, sonicated and centrifuged at 13,500xg for 10 minutes at 4°C. Lysate were run on a 10% polyacrylamide gel with equivalent volumes loaded of the soluble and insoluble fractions per each sample. For Western Blotting analysis the following antibodies were used: mouse anti-HA (Cell Signaling Cat. No 2367) and IRDye 680 anti-mouse secondary antibody for detection (Licor); rabbit GAPDH and rabbit Lamin (used as controls) and 800 anti-rabbit secondary antibody (Licor).

3.8 Immunoprecipitations

V5-PFN1 transfected HEK293 cells were lysed at 24 hours with RIPA buffer and immunoprecitated with 1µg of anti-V5 antibody overnight at 4°C followed by incubation with Dynabeads Protein G for 1 hour. The protein-bead complexes were washed four times with RIPA buffer, eluted by boiling at 95°C for 5 minutes and then subject to western blot analysis to detect V5-PFN1 and actin. The following antibodies were used: mouse anti-V5 (1:5000, Invitrogen), mouse anti-beta actin (1:1000, Sigma), goat anti-mouse 800CW (1:10000, LICOR).

3.9 Tubulin Dimerization Assay

In vitro translation of HA-TUBA4A constructs was performed using the TNT T7 Coupled Reticulocyte Lysate system (Promega). For non-denaturing gel electrophoresis, the products were diluted 1:1 in sample buffer (160 mM MES pH 6.86, 2mM MgCl2, 2 mM EGTA, 1 mM GTP, 20% Glycerol), immediately loaded onto a non denaturing gel (4.5% polyacrylamide in 80mM MES pH 6.86, 1mM MgCl2, 1mM EGTA, 1mM GTP) and run at 95 volts for 1.5 hours. For SDS-gel electrophoresis, the reaction products were diluted in 2X Laemmli buffer, boiled for 5 minutes and loaded onto a 4-20% Mini-PROTEAN TGX gel (Bio-Rad). Proteins were transferred to a 0.45 µm nitrocellulose membrane for 10 minutes at 1.2 amp/25 volts on the Trans-Blot Turbo Transfer System (Bio-Rad). HA-TUBA4A was detected via chemiluminescence using mouse HA antibodies (MMS-101R-200, Covance), goat anti-mouse HRP conjugate (sc-2005, Santa Cruz) and SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific).

4. RESULTS

We performed an exome-sequencing study to identify novel causative genes in familial and sporadic ALS and applied specific analysis strategies modeled on different study designs.

4.1 Exome sequencing of ALS families identifies mutations in *PFN1* gene

In the first phase of our study, we investigated 2 large dominant ALS families of Caucasian (Family #1, Figure 5A) and Sephardic Jewish (Family #2, Figure 5B) origin, negative for mutations in known ALS-associated genes *(SOD1, C9orf72, TARDBP, FUS, ANG*). A combined strategy of whole-exome sequencing and linkage analysis was employed to identify the underlying mutations. Exome-sequencing was performed on the two affected members from each family with maximum genetic distance (III:8 and III:12 in Family#1, II:4 and II:11 in Family#2). A high level of coverage (>150X) was achieved with an average of 1.1×10^{10} and 2.3×10^{10} base pairs sequenced for Families #1 and #2, respectively. After filtering to exclude variants reported in public databases (dbSNP132, 1000 Genomes Project, NHLBI ESP Exome Variant Server) and combining data with linkage analysis, we observed 2 and 7 heterozygous coding variants in Family#1 and Family#2 respectively, shared by both affected members in each family (Table 5). Validation by Sanger sequencing in the probands and proof of segregation of the variants in additional affected members reduced the list of candidates to two variants in Family #1 and three variants in Family #2 (Table 5). Interestingly, both families harbored different mutations (C71G and M114T) within the same gene *PFN1,* profilin-1 (Table 6).

Table 5. Filtering steps applied to variant detected by WES in FALS pedigree. The table reports the number of variants left after each filtering steps. The regions with LOD score>zero were calculated on linkage analysis of FALS pedigrees.

FILTER	Family #1	Family #2
Total variants	282.782	382.751
Coding variants	29.777	42.661
Overlapping variants	9.045	10.669
Novel variants	18	178
Heterozygous/Non-synonymous variants	10	135
Lod score>0	2	7
Confirmed by Sanger sequencing	2	6
Mendelian segregation	2	3
Totale	**2**	**3**

Table 6. Candidate causal genes for familial ALS after filtering steps.

	Gene	Description	Mutation	SIFT
Family #1	PFN1	Profilin-1	C71G	damaging
	XPOT	tRNA exportin	V139A	tolerated
Family #2	FMO2	Dimethylaniline mono-oxygenase 2	T390I	tolerated
	KIF1C	Kinsesin-like protein KIF1C	I118L	tolerated
	PFN1	Profilin-1	M114T	damaging

PFN1, located on chromosome 17p13.2, encodes for a 140 amino acid protein regulating monomeric (G)-actin conversion to filamentous (F)-actin filamentous [87]. Based on these data, we postulated that mutations in the *PFN1* gene can cause familial ALS.

To further investigate the causative role of *PFN1* gene in ALS, we sequenced the entire coding region in a large cohort of 272 index FALS cases and identified five additional cases harboring alterations in *PFN1.* The C71G missense mutation, originally identified in Family #1, was detected in two additional families (Family #3 and Family #5, Figure 6) and segregated with the disease in Family #3, in which three other affected family members were available for testing. The other mutations included: a M114T variant, identified in two affected siblings of an ALS family of Italian origin (Family #4, Figure 6); a consecutive base pair change (AA to GT) resulting in a E117G mutation and a G118V mutation.

Given that many familial ALS-associated genes account also for a small percentage of sporadic cases, we investigated the genetic contribution of *PFN1* also in sporadic ALS. Sequencing of the entire coding region in a large cohort of 1984 SALS resulted in the identification of the variant p.E117G in three additional cases.

In total, we identified 4 different *PFN1* mutations in 7/274 FALS cases (2.6%) and the E117G variant in 3/1984 SALS (Table 7): all mutated cases but one had spinal onset ALS and none manifested signs of dementia. The clinical phenotype of the patients carrying *PFN1* mutations is reported in Table 8. Considering the great pathological and genetic overlap between ALS and FTD, we extended the analysis of *PFN1* gene to 203 FTD patients (of whom 30 represented familial cases) but we failed to identify any mutation (Table 7).

Three of the ALS-related mutations (C71G, M114T, G118V) were not found in a total of 9072 controls, including 2601 individuals analyzed in-house by genotyping or Sanger sequencing and 6471 controls available through 1000 Genomes Project database and NHLBI ESP Exome Variant Server (Table 7). Conversely, the E117G variant was observed in 3/9072 control subjects, with an allele frequency of 1.77×10^{-3} in ALS vs 3.03×10^{-4} in controls.

Figure 6. Pedigrees of families with additional PFN1 mutations. Pedigrees of Families #3-7, harbouring PFN1 mutations, are shown. The gender of the pedigree members is obscured to protect privacy. Mutant alleles are indicated by mut, whereas wild-type alleles are indicated by wt.

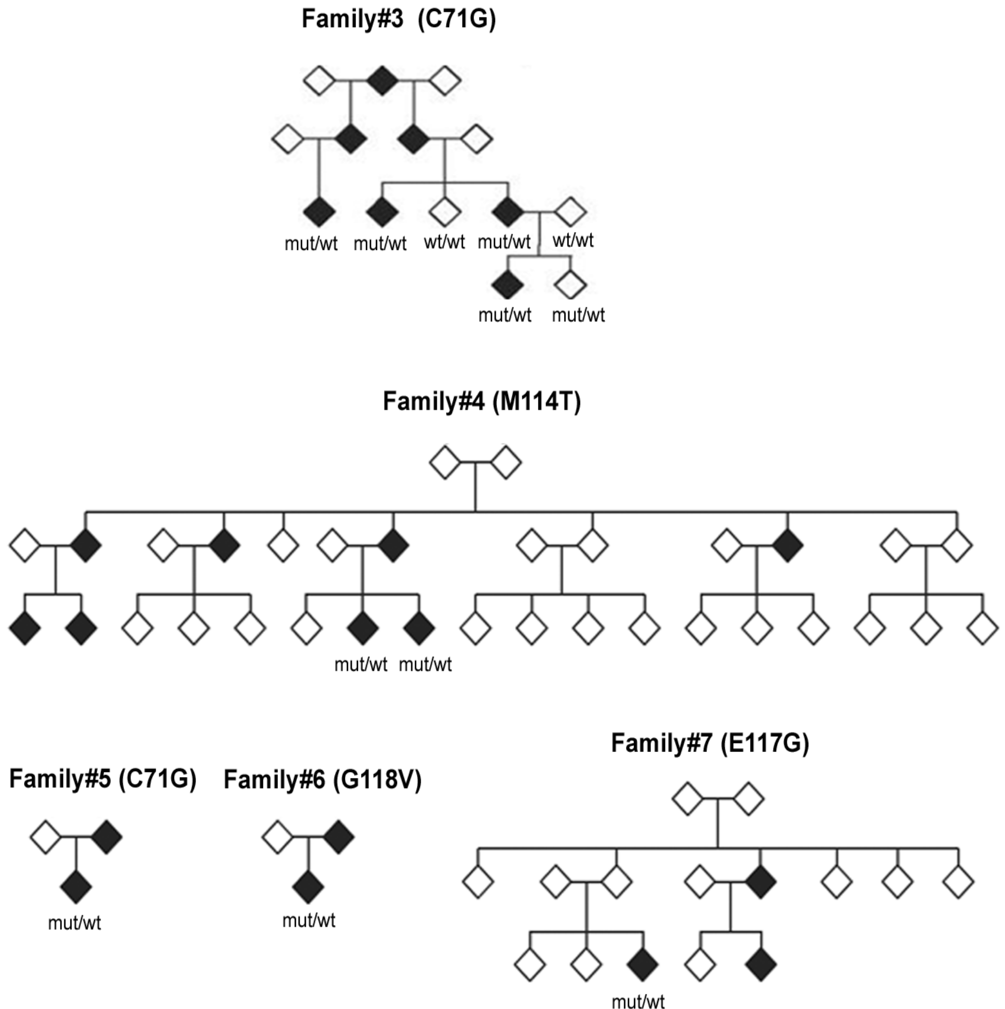


Table 7. Results of *PFN1* mutational screening in FALS, SALS, FTD patients and controls.

Mutation	Patients			Controls			
	FALS	SALS	FTD	1000 Genome	ESP	Genotyping/ Sequencing	Total
C71G	2/274	0/1984	0/203	0/1092	0/5379	0/2601	0/9072
M114T	1/274	0/1984	0/203	0/1092	0/5379	0/2601	0/9072
E117G	1/274	3/1984	0/203	0/1092	2/5379	1/2601	3/9072
G118V	1/274	0/1984	0/203	0/1092	0/5379	0/2601	0/9072

Table 8. Clinical phenotype of ALS patients harbouring *PFN1* mutations

Patients	Mutation	Age of Onset (yrs)	Site of Onset	Dementia	No. Pts
Family # 3	C71G	41.1 +/- 4.3	spinal	no	4
Family # 1	C71G	50.0 +/- 6.6	spinal	no	4
Family # 5	C71G	44.0	spinal	no	1
Family # 2	M114T	41.9 +/- 5.3	spinal	no	7
Family # 4	M114T	52.0 +/- 13.1	spinal	no	3
Family # 7	E117G	40.3	spinal	no	1
Family # 6	G118V	43.0	spinal	no	1
SALS# 1	E117G	63.0	spinal	no	1
SALS# 2	E117G	33.9	spinal	no	1
SALS# 3	E117G	*72.0*	bulbar	no	1

4.2 PFN1 mutants are aggregation prone and show alterations in actin binding.

In light of these genetic findings, a functional characterization of the identified ALS-associated variants was conducted to investigate their potential pathogenicity in ALS.

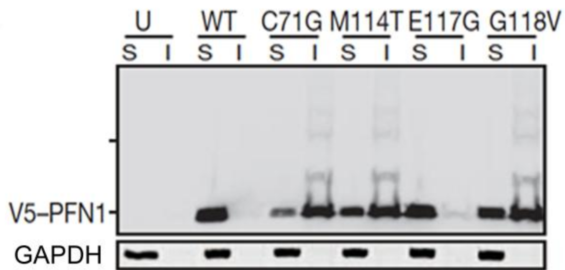
Since misfolding and aggregation are key features demonstrated for many ALS-linked proteins (including FUS, SOD1, OPTN, UBQLN2, HNRNPA2/B1, VCP), we started to investigate whether mutations reduce profilin-1 solubility, inducing the formation of insoluble aggregates.

To test this hypothesis we performed a Western-blot analysis of NP-40-soluble and insoluble fractions of N2A cells transfected with wild-type (WT) or mutant PFN1. Compared to WT PFN1, which was present predominantly in the soluble fraction (Figure 7a), *PFN1* mutants were considerably detected in the insoluble fractions, where several higher molecular weight species (indicative of SDS-resistant oligomers) were also observed. Only the E117G mutant displayed a pattern more similar to wild-type PFN1 with most of the expressed protein in the soluble fraction. Impairment of ubiquitin-proteosome system induced by treatment with MG132 resulted in increased levels of PFN1 mutants in insoluble fraction (Figure 7b).

The propensity of PFN1 mutant to form insoluble ubiquitinated aggregates was further investigated by immunostaing profilin-1 in transfected NA2 cells and primary motor neurons. In N2A cells wild-type PFN1 exhibited a diffuse cytoplasmic localization but, in contrast, 15-61% of mutant expressing cells displayed ubiquitinated cytoplasmic aggregates containing profilin-1 (Figure 8a-b). Similar ubiquitinated aggregates were observed in primary motor neurons (PMNs) expressing the C71G, M114T, and G118V PFN1 mutants, while aggregates were not observed in cells expressing WT and E117G construct (Figure 8c). Co-aggregation of TDP-43 with mutant PFN1 was observed (Figure 8d).

Figure 7. Different solubility of PFN1 mutants. Western blot analysis of NP-40 soluble and insoluble fractions in transfected N2A cells not (**a**) or treated with MG132 (**b**).

a



b

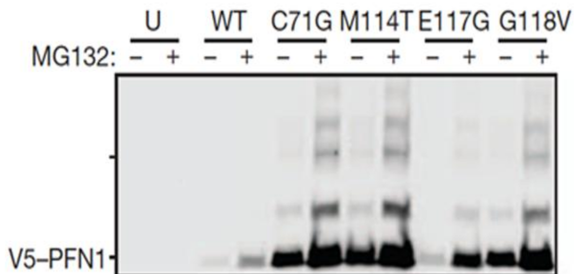
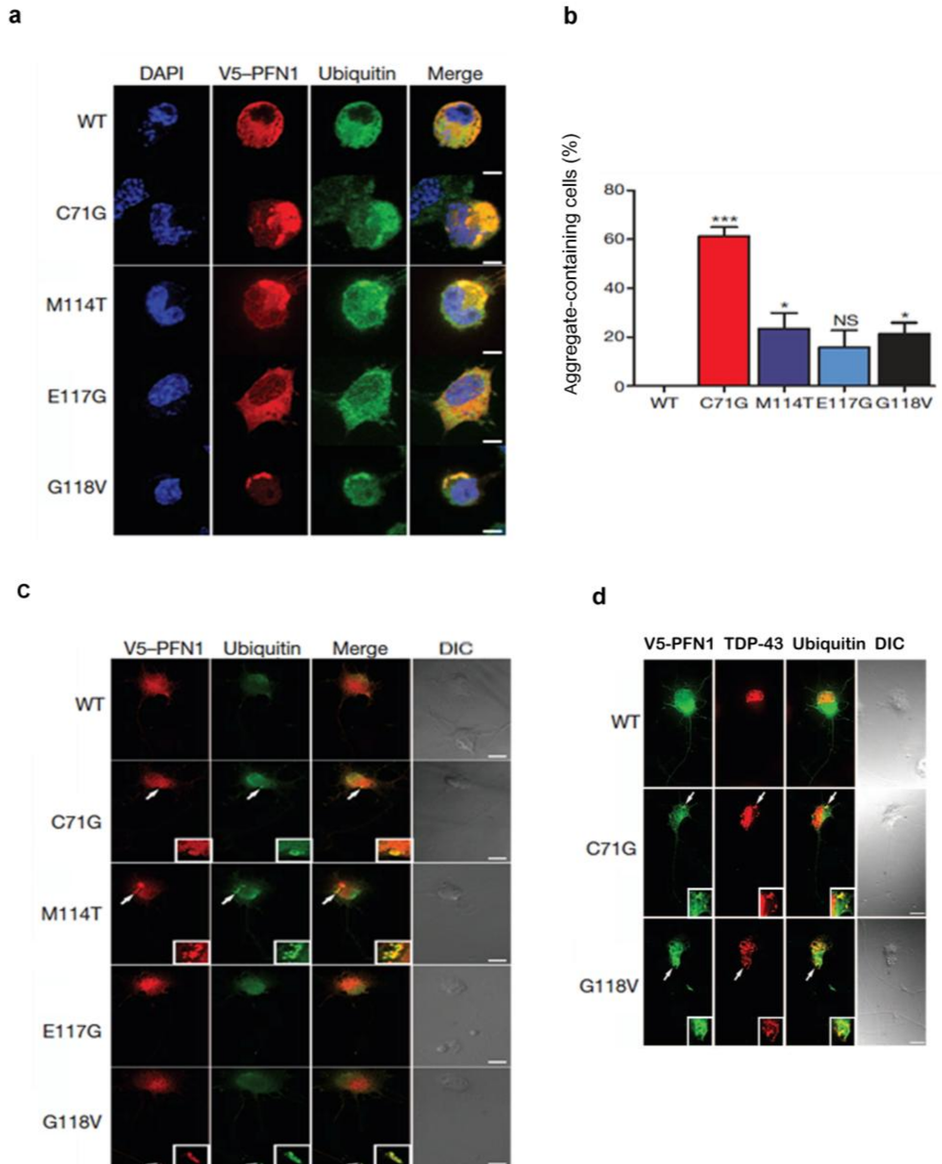


Figure 8. Mutant PFN1 produces ubiquitinated insoluble aggregates. (a-b) Transfected N2A cells were stained with V5 and ubiquitin antibodies and cell displaying insoluble aggregates were counted and analyzed using one-way ANOVA testing with Dunnett's multiple test comparison (n=127-135 transfected cells from 3 independent experiments). *P<0.05, ***P<0.001, n.s. P>0.05. Error bars indicate SEM. (**c,d)** Transfected PMNs stained with V5 and TDP-43 antibodies. Scale bars: 5 μm (**a**), 10 μm (**c, d**)



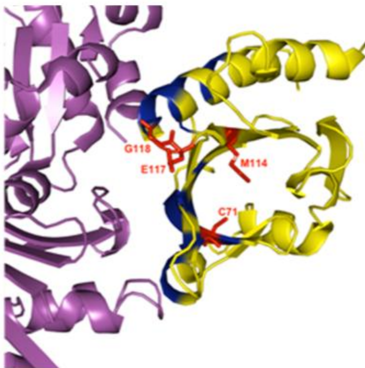
PFN1 plays an important role in actin polymerization by catalyzing ADP-to-ATP exchange on G-actin and adding ATP-G-actin to the barbed ends of actin filaments. Therefore, we evaluated the possible pathological effects of mutated PFN1 on actin polymerization. Since structural analysis of the profilin1-actin complex showed that all ALS-linked mutations lie in close proximity to the actin binding domain of profilin-1 (Figure 9a), we conducted a co-immunoprecipitation assay to test the actin binding ability of each mutant. The synthetic H120E PFN1 mutant, known to abolish the binding to actin, was used in experiments as positive control [88]. After performing co-immunoprecipitation experiments using lysates of transfected HEK293 cells and Western Blot analysis, we observed that the C71G, M114T, G118V and H120E mutants displayed reduced levels of bound actin compared to WT profilin-1(Figure 9b). The E117G mutant did not display a reduction in actin binding compared to the wild-type counterpart.

Since previous reports revealed that expression of profilin-1 mutants defective in actin binding decreases neurite length and strongly inhibits filopodia formation [78], the effect of ALS-linked PFN1 mutants on neurite outgrowth was evaluated by measuring axonal length in transfected PMNs. Three mutants (C71G, M114T, and G118V) induced a significant decrease in axon outgrowth, similar to the reduction observed for the H120E construct (Figure 10). Axon outgrowth inhibition was observed also with the E117G mutant, although not reaching statistical significance.

Together, these results suggest that ALS-associated PFN1 mutations display several functional defects linked to ALS pathogenesis: an increased aggregation propensity in cell cultures and an altered actin-binding ability with inhibitory effect on axon outgrowth. According to experimental observations, p.E117G is likely variant with reduced pathogenic potential, since it exhibits only moderate aggregation propensity and does not affect actin binding and polymerization.

Figure 9. PFN1 mutations display reduced acting-binding ability. (a) PFN1-actin interaction region using the PyMOL Molecular Graphics System (v. 1.4)**.** Magenta:actin; Yellow:PFN1; Green:actin-binding PFN1 residues; Red: ALS-linked mutated PFN1 residues; **(b)** Cell lysates of transfected HEK293 cells were immunoprecipitated with anti-V5 antibody and then analyzed by immunoblotting with antibodies as indicated**.**

a



b

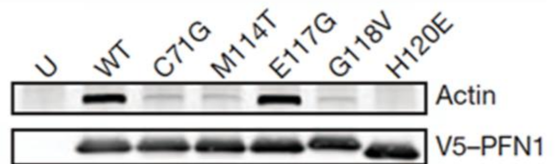
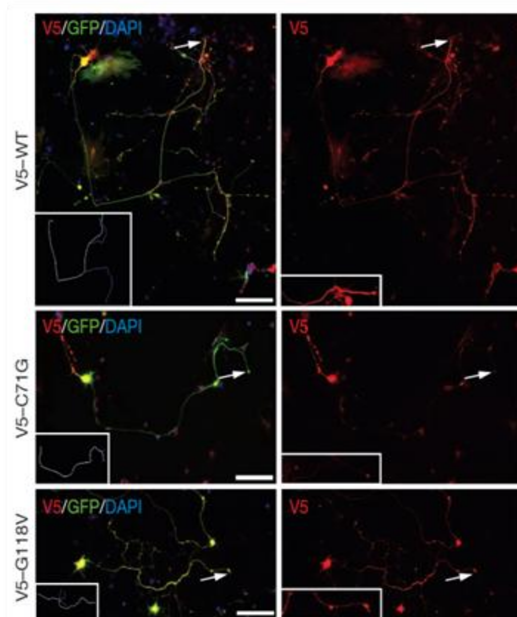
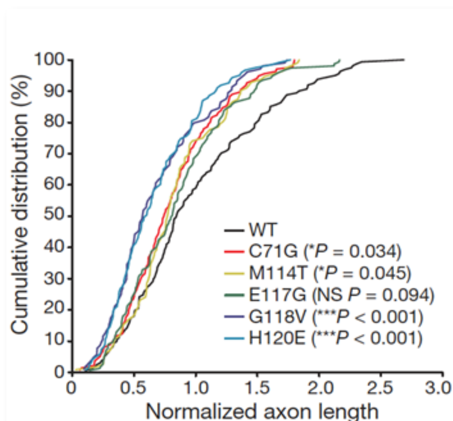


Figure 10. Mutant PFN1 inhibits axon outgrowth. PMNs transfected with wild-type or mutant V5-PFN1 and a GFP expressing construct were stained to detect V5-PFN1. The axon tip, indicated by arrows, is enlarged in the inset in right panel. Cumulative distribution of axon lengths relative to the mean of wild-type PFN1 transfected cells was plotted.P values are given in the legend (n=104-161 cells from 4 independent experiments). Scale bar: 100 μm.





4.3 Exome sequencing of unrelated FALS indentifies TUBA4A as candidate gene

Because of the difficulty in collecting an adequate number of multi-generational FALS informative pedigrees, we also sequenced the exomes of 363 index familial cases (one affected sample per family) with unknown genetic cause and included 9 FALS with known causative mutations (6 mutated for *SOD1* and 3 for *VCP*) as positive controls. An average of 3.2×10^{9} target bases were sequenced per sample to an average depth of 90.4X. Filtering steps led to the elimination of variants not passing quality control criteria (QD<5.0, HRun>3, MQ<40.0, HaplotypeScore>13.0, FS>60.0, , MQRankSum< -12.5, ReadPosRankSum< -8.0) and to the exclusion of samples with low call rate. An exome-wide rare variant analysis, comparing the total number of gene variants seen among unrelated cases and controls, was performed to identify potential candidate genes. The controls were represented by 31 healthy samples sequenced in house and by 4300 European Americans individuals from the NHLBI's Exome Variant Server (EVS). The analysis was performed for 12,495 genes fulfilling all quality control filters and restricted only to variants with a MAF <0.0004 in EVS and predicted to be damaging using PolyPhen-2 or resulting in a stop gain/loss.

As shown in Figure 11, *SOD1* gene (6 cases [1.6%] vs 0 controls, $P=5.5 \times 10^{-8}$, $P_{corrected}=6.2 \times 10^{-4}$) ranked first and *VCP* (3 cases [0.8%] vs 0 controls, $P=1.1 \times 10^{-4}$, $P_{corrected}=0.73$) ranked tenth among all genes confirming the validity of our approach. After *SOD1*, the top hit was *TUBA4A*, encoding for Tubulin, Alpha 4A (4 cases [1.1%] vs 0 controls, $P=9.1 \times 10^{-6}$, $P_{corrected}=0.09$). The gene was well covered by EVS confirming that the P-value was not inflated due to poor control sequencing. Although *TUBA4A* was the most consistent hit, the analysis identified other possible candidate genes for FALS, including *MATR3* which was recently recognized as FALS-causative gene by exome-sequencing [89].

Five different *TUBA4A* variants were identified by exome-sequencing in 363 FALS: all mutations were confirmed by Sanger sequencing in relevant cases and were located in the last exon of the gene, affecting evolutionary highly conserved residues (Figure 12). Then nonsynonymous substitutions included 4 missense mutations (R215C, R320C, R320H, A383T) and 1 nonsense mutation (W407X) removing the last 41 aminoacids of the C-terminal region. Unfortunately, no DNA was available from additional affected relatives to test segregation. None of these mutations were observed in the 4300 European American EVS controls, in which only three non-synonymous changes (not predicted to be damaging) were observed.

To further investigate the role of *TUBA4A* as an ALS causative gene, we sequenced an independent replication panel of 272 index FALS cases and identified two additional variants (T145P, K430N) in two FALS: the mutation T145P was predicted to be probably damaging by PolyPhen-2 (score=0.998) and segregated in the affected parent of the proband; the other variant K430N was not observed in the affected first cousin of the proband, representing probably a neutral rare polymorphism. By extending our mutational analysis to 1355 SALS, we identified another mutation (G43V) in one sample: this variant was predicted to be benign according to PolyPhen-2.

Five FALS-related TUBA4A mutations (T145P, R215C, R320C/H, W407X) were not found in a total of 13,063 control samples, including 4500 European Americans and 2200 African Americans individuals available from EVS, 1053 individuals from 1000 Genome Project and 5510 European Americans subjects sequenced in house (Table 9). However, the G43V and A383T substitutions were observed in a single Italian and an African American control, respectively.

Figure 11. Rare variant analysis identifies TUBA4A mutations in FALS. Manhattan plot displaying permutation-based corrected P values generated by a rare variant analysis of FALS. The dotted line represents a Pcorrected=0.05.

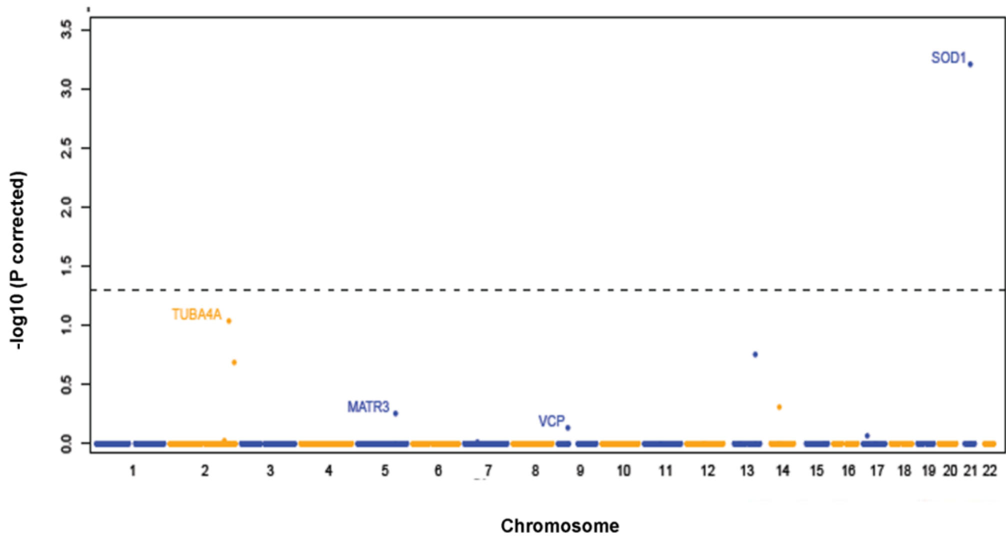


Figure 12. Evolutionary conservation of TUBA4A residues affected by mutation. Aminoacid alignment of TUBA4A protein showing conservation of mutates residues in different phyla. The mutated residue is in red.

	G43V	T145P	R215C	R320C/H	A383T	W407X	K430N
Mutation	-----V-----	-----P-----	-----C-----	-----C----- -----H-----	-----T-----	-----X-----	-----N-----
Human	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Chimp	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Rhesus	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Mouse	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Rat	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Rabbit	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Cow	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Dog	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Elephant	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM
Zebrafish	SDKTIGGGDDS	SFGGGTGSGFT	YDICRRNLDIE	CCLLYRGDVVP	LSNTTAIAEAW	RAFVHWYVGEG	VEEYDKELAAM

Table 9. Analysis of TUBA4A mutations in controls. To confirm that the newly identified *TUBA4A* variants represent causal mutations, 5510 Caucasian controls were directly sequenced and each variant was interrogated in the 1000 Genomes Project database and the NHLBI ESP Exome Variant Server.

Mutation	EVS-EA	EVS-AA	1000 Genome Project	Internal controls	Total
G43V	0/4300	0/2200	0/1053	1/5510	1/13063
T145P	0/4300	0/2200	0/1053	0/5510	0/13063
R215C	0/4300	0/2200	0/1053	0/5510	0/13063
R320C	0/4300	0/2200	0/1053	0/5510	0/13063
R320H	0/4300	0/2200	0/1053	0/5510	0/13063
A383T	0/4300	1/2200	0/1053	0/5510	1/13063
W407X	0/4300	0/2200	0/1053	0/5510	0/13063
K430N	0/4300	0/2200	0/1053	0/5510	0/13063

EVS (Exome Variant Server, NHLBI GO Exome Sequencing Project)

EA (European-Americans), AA (African-Americans)

The clinical features of ALS patients carrying TUBA4A mutations are shown in Table 10. All the patients had spinal onset ALS with upper and lower motor neuron signs. Two cases developed a cognitive decline of frontal type, consistent with a diagnosis of frontotemporal dementia (FTD) and another reported a first-degree relative with FTD, suggesting that TUBA4A mutations may also contribute to FTD pathogenesis

Table 10. Clinical characteristics of ALS patients with TUBA4A mutations. Disease duration is calculated in months to the last follow-up.

Mutation	Family history	Other ND family history	Age of onset	Site of onset	Disease durarion (months)	Dementia
G43V	SALS	dementia	52	spinal	45	FTD
T145P	FALS	dementia	48	spinal	69	no
R215C	FALS	dementia	78	spinal	n/a	FTD
R320C	FALS	no	64	spinal	12	no
R320H	FALS	no	41	spinal	33	no
A383T	FALS	no	71	spinal	18	no
W407X	FALS	no	66	spinal	36	no
K430N	FALS	no	64	spinal	42	no

ND (neurodegenerative disorders)

4.4 TUBA4A mutants affect microtubule polymerization

Based on these genetic finding, we explored the functional consequences of mutations in *TUBA4A*. Since several causative ALS proteins, including profilin-1, have been shown to form insoluble inclusions in cell culture, we assessed the tendency of the TUBA4A mutants to aggregate.

After transfecting HEK293 cells with WT and mutant HA-tagged TUBA4A constructs, we performed a Western Blot analysis of fractions obtained by differential solubilization in four different buffers (PBS, Triton X-100, SDS, urea). We observed that the majority of W407X mutant protein was insoluble in Triton X-100 being detected prevalently in the SDS and urea fractions. The other mutants (G43V,R215C, R320C/H, A383T) displayed no significant differences in solubility compared to WT (Figure 13).

Additionally, the formation of insoluble aggregates was further investigated by immunostaining HEK293 cells and primary motor neurons (PMNs) transfected with TUBA4A constructs. Consistent with the results obtained in soluble/insoluble fractionation experiments, we observed that the W407X mutant did not incorporate into the microtubule network and formed small ubiquitinated cytoplasmic inclusions in ~85% of transfected HEK293 cells and in ~40% of transfected PMNs (Figure 14). The other TUBA4A mutants formed cytoplasmic inclusions in only 10-30% of transfected HEK293 cells (Figure 14) while no aggregation was observed in PMNs. However, subtle alterations in TUBA4A cytoplasmic distribution were observed, such as a more diffuse cytoplasmic staining compared to the wild-type protein, which was mainly distributed within the microtubule network.

Figure 13. Protein fractionation of TUBA4A mutations. **(A)** HEK cells were transfected with constructs expressing wild-type or mutant TUBA4A. Cellular lysates were subject to sequential precipitations and resuspensions using the buffers indicated and immunoblotted with anti-HA anitibody. **(B)** Expression levels were quantified and the percent of the total HA-TUBA4A protein was calculated. Bars represent mean and SEM (one-way ANOVA and Dunnett's *post hoc* test, n=3, ** *P*<0.01).

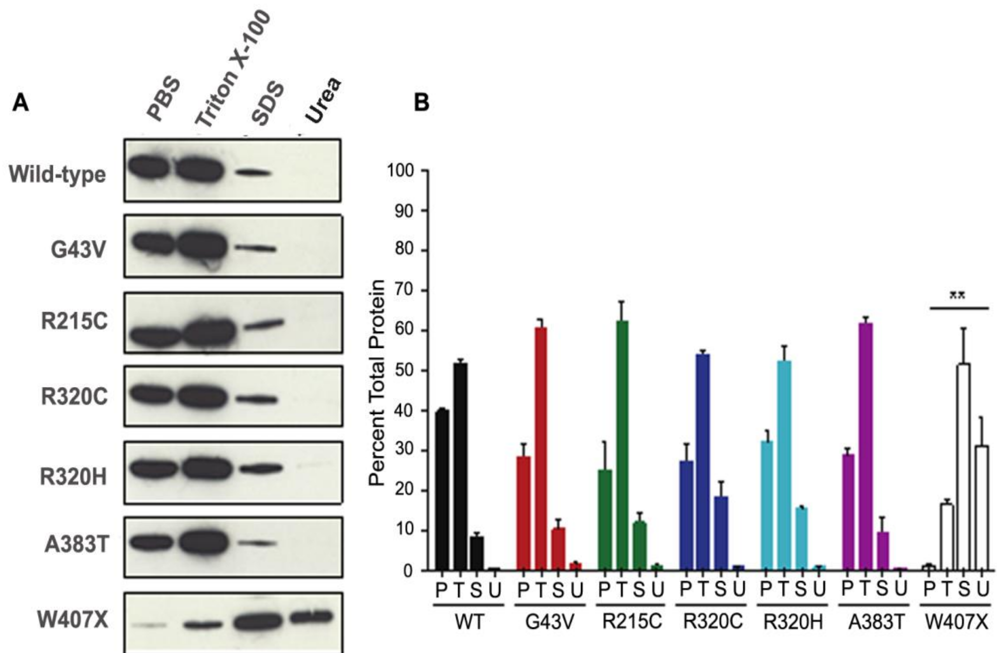
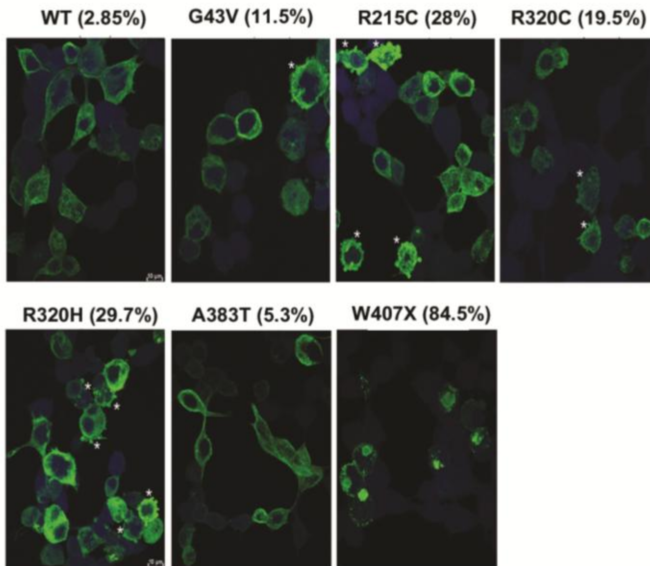
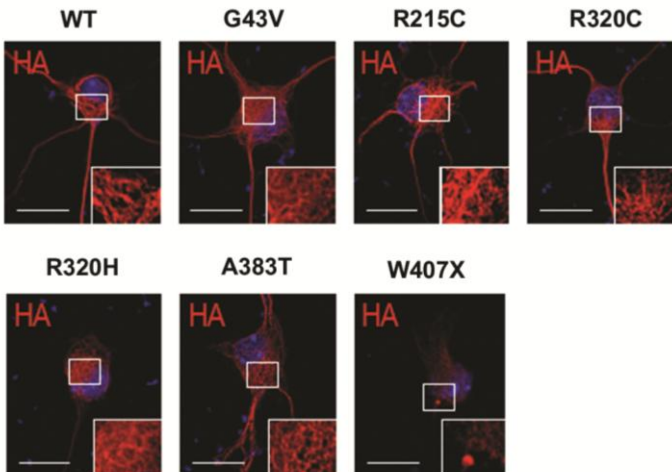


Figure 14. Mutant TUBA4A W407X forms ubiquitinated agreggates. (a) HEK293 cells transfected with wild-type or mutant HA-TUBA4A constructs were immunostained for HA. TUBA4A-positive aggregates were particularly abudant in the cells trasnfected with W407X mutant. **(b)** PMNs transfected with HA-tagged TUBA4A constructs. The presence of TUBA4A and ubiquitin postive aggregates was oberved only for the mutant W407X.





The protein encoded by *TUBA4A* belongs to the tubulin superfamily and is a major constituent of microtubules, which are involved in many cellular processes (including mitosis, cell motility, intracellular transport, secretion, maintenance of cell shape and cell polarization). In particular microtubules are composed by alpha- and beta-tubulin heterodimer subunits assembled into linear protofilaments. Therefore, we started to investigate the ability of TUBA4A mutants to be efficiently incorporated in alpha/beta tubulin dimers by performing an *in vitro* translation in rabbit reticulocyte lysate, which contains all the components needed to form tubulin dimers. The analysis of the reaction products in denaturating conditions (SDS-PAGE) showed that the translation efficiency of all the mutants was similar to the wild-type (Figure 15). The same reaction products were then analyzed by non-denaturing gel electrophoresis to measure incorporation of the translated recombinant protein into tubulin heterodimers. The mutant protein W407X, which deletes the C-terminal region interacting with β-tubulin, did not form heterodimers at all, while two other mutants, A383T and R320, displayed an impaired dimer formation compared to the WT protein (Figure 15). The other mutants did not differ significantly from the control.

The functional characterization of TUBA4A mutants was extended to test their capacity *in vivo* to incorporate into the microtubules of transfected primary glial cells, which have a large cytoplasm and a relatively long cell cycle that result in a extensive and stable microtubule network. According to the distribution and microtubule incorporation of HA-tagged TUBA4A, the cells were classified into four categories (normal, mild, moderate, severe). Contrary to the wild-type protein, the R320C/H, A383T and R215C mutants all displayed a significantly different distribution and a defective incorporation into microtubules (Figure 16). In particular, we found that the W407X mutant failed to be incorporated in microtubules, showing a diffuse

distribution throughout the cytoplasm and the presence of small aggregate-like inclusions. No difference was detected for the G43V mutant.

Together, these results suggest that ALS-associated TUBA4A mutations exhibit several defects, from the aggregation propensity of the truncated protein W407X to the mild or more severe deficiencies showed also by the other mutants in heterodimerization and microtubules formation.

Figure 15. Effects of TUBA4A mutations on tubulin heterodimerization. Non-denaturing (top) and SDS-gel electrophoresis (bottom) of HA-TUBA4A constructs after *in vitro* translation. The intensity of the bands representing the α/β tubulin dimers was quantified from 6 independent experiments. Bars represent mean and SEM (one-way ANOVA and Dunnett's *post hoc* test, * $p<0.05$, *** $p<0.001$).

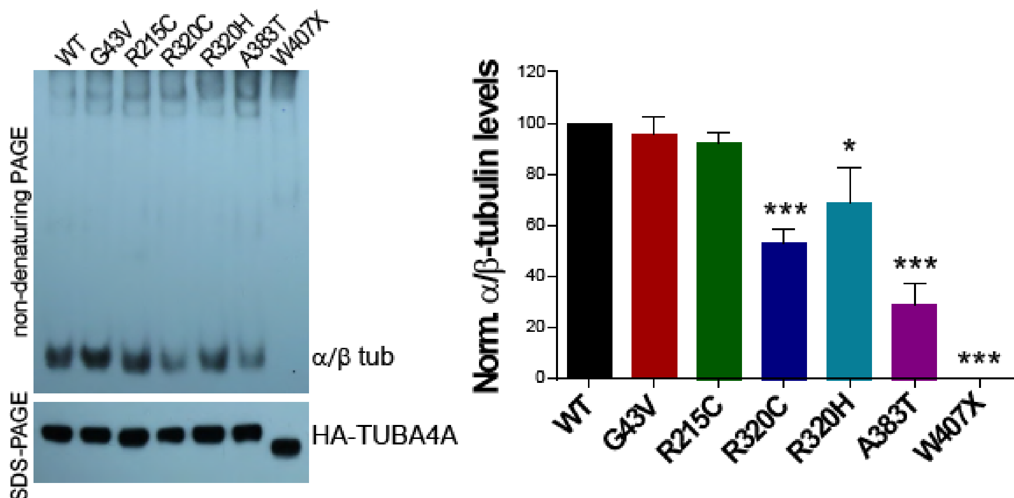
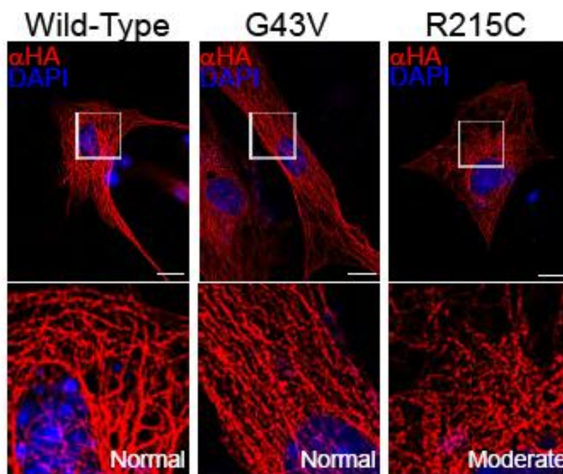
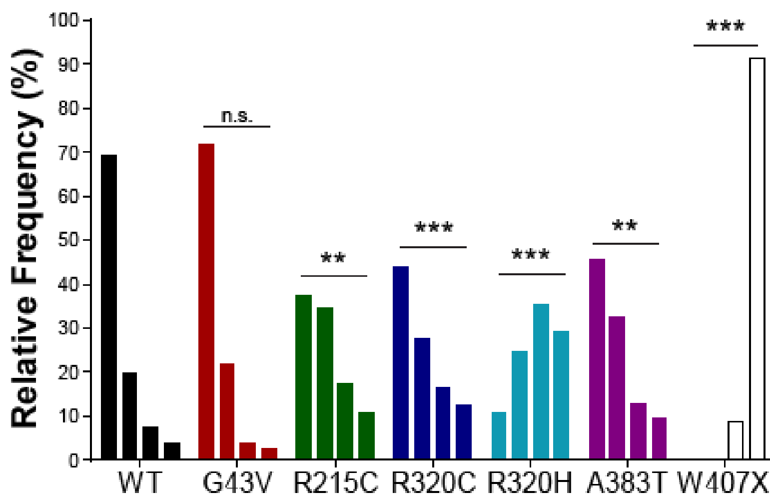


Figure 16. Mutant TUBA4A alters microtubule polymerization Primary glial cells were transfected with the HA-TUBA4A constructs and fixed 4 days after transfection. The distribution of the HA-TUBA4A protein was classified as normal, mild (subtle diffuse staining), moderate (significant diffuse staining), or severe (no visible HA-positive microtubules). f, The frequency of each phenotype was quantified and compared between the wild-type and mutant proteins (Kolmogorov-Smirnov test; n=100-130 cells per condition from 4 independent experiments, * $p<0.05$, ** $p<0.01$, *** $p<0.001$). Red, HA; green, ubiquitin; blue, DAPI. White boxes indicate the regions enlarged in the insets. Scale bars: 10μm





4.5 Identification of *de novo* mutations using a trio-based exome sequencing approach

To identify further genetic factors contributing to ALS, we assessed the role of *de novo* mutations by applying the exome-sequencing approach to ALS trios (the proband and both unaffected parents). We selected 32 ALS trios of Italian origin, each patient with a diagnosis of sporadic ALS according to the EI Escorial revised criteria [24] and testing negative for mutations in the main ALS-causative genes. Target cature and exome-sequencing were performed on a total of 96 individuals achieving an average of 3.2 x 10^9 base pairs sequenced per sample to an average depth of 90.4X.

In the filtering process, all the inherited genomic variants were excluded and *de novo* mutations (DNMs), identified as sites where both parents were homozygous for the reference sequence and the offspring was heterozygous, were reteined. DNMs were then filtered against dbSNPv135, the NHLBI Exome Sequencing Project and the 1000 Genomes Project to select only rare or novel variants, on the assumption that disease causing variants are unlikely to be present in controls.

By performing a final validation with Sanger sequencing on the remaining candidates, we identified 25 *de novo* non-synonymous variants in 16 ALS trios and no DNMs in the remaining 16 ALS trios. The overall de novo mutation rate observed in our ALS trios was 0.78, consistent with those reported in recent WGS studies [78].

Among the DNMs identified, 24 were missense mutations and only 1 was a nonsense substitution. The list of all DNMs identified in ALS trios is reported in Table 11. We did not identify DNMs in previousily known ALS-associated genes.

Two variants (G358S in *C10orf90* and R709W in *FGD5*) were found to be annotated in the NHLBI ESP database with a minor allele frequency (MAF) <0.0001, while the remaining DNMs were novel. Sixteen of the 24 missense variants were predicted to be potentially damaging according to Poly-Phen-2 and SIFT score. However, none of the genes contained multiple DMNs among different trios or represented genes identified in other neurodegerative diseases or similar phenotypes.
To narrow the list of candidate genes to prioritize, we therefore assessed the possible pathogenic role of DNM-containing genes by searching for additional mutations in FALS cases. For this analysis we used the available exome sequencing data from 220 Italian and US FALS patients. Interestingly, 30 additional novel mutations (all missense) in 17 of our 25 DNMs-containing genes were identified in FALS (Table 12). However, none of the original DNMs was observed in our FALS cohort.

Since none of DNMs containing genes has been previously associated with ALS, we applied different bioinformatic tools to investigate if the protein products of these genes interact with each other or if they are connected to a set of known ALS-associated genes. A complet list of genes that have been associated to ALS as causative or risk factors was extracted by ALSOD database (available at http://alsod.iop.kcl.ac.uk/).
Initially, protein-to-protein interaction assessment was conducted using the Disease Association Protein-Protein Link Evaluator (DAPPLE) tool (http://www.broadinstitute.org/mpg/dapple/dapple.php) on the 25 genes containing DNMs. In silico analysis showed one significant direct connectivity among two of our candidate DNMs genes, RRM33 and ASCC3, with a corrected p-value less than 0.05 (Figure 17). An indirect connectivity emerged also for 5 other genes (TJP3, SH2D3A, VPS11, RRP1B, LTV1). When known ALS-related genes were included in the analysis, the number of candidate proteins partecipating into the neteork

increased: RRM33 and ASCC33 showed direct connectivity with ATXN2 (Figure 18) and other connections emerged between the candidate gene EPHA2 and EPHA4, identified as a modifier gene in ALS [90], and TJP3 with GRN (Figure 18).

We then investigated if candidate DNMs genes belonged to related functional pathways. Functional classification analysis was performed with David in our set of genes impacted by DNMs, revealing that some functional categories resulted enriched. In particular, 6 (TRPM4, SLC6A17, SLC9A5, CLCN2, VPS11 and c9orf72) out of 25 genes resulted implicated in transport activity, while 3 genes (ARHGEF18, FDG5, SH2D3A) showed guanyl-nucleotide exchange factor activity and GTPase regulator activity (Enrichment Score 1.36) thus having similar functions to alsin, which activates members of the Ras superfamily of GTPases and has been identified as causative of ALS. To enhance the power of bioinformatic analysis, 26 genes identified in 40 ALS trios from a previously published study [78] were included in the analysis for a total of 51 imputs. From this analysis two other genes, STARD13 and KIF13A, enriched the functional categories "GTPase regulator activity" and "transport", respectively. Interestingly, 2 (ASCC3 and ZFP64) of our genes together with other 5 genes from the previous study were annotated in the same cluster, being all involved in transcription regulation.

Table 11. Candidate DNMs identified in ALS trios. All variants were validated by Sanger sequencing in relevant trios**.** The effect of missense mutations was tested by PolyPhen-2 and SIFT**.**

gene	Chr	aa change	Mutation	polyPhen	ESP MAF
ADAM33	chr20	R734X	stop-gained	N/A	0
ARHGEF18	chr19	A1015T	Missense	benign	0
ASCC3	chr6	D1912Y	missense	probably damaging	0
C10orf90	chr10	G358S	missense	possible damaging	0,000077
C8B	chr1	C110F	missense	probably damaging	0
CLCN2	chr3	R290Q	missense	probably damaging	0
EPHA2	chr1	Q368R	missense	benign	0
FGD5	chr3	R709W	missense	probably damaging	0,000084
LTBP4	chr19	A1392T	missense	probably damaging	0
LTV1	chr6	E296K	missense	possible damaging	0
MFSD12	chr19	G28D	missense	probably damaging	0
MRGPRG	chr11	V101L	missense	benign	0
ODZ3	chr4	P219A	missense	benign	0
PAPPA	chr9	V353M	missense	probably damaging	0
PTCRA	chr6	G280V	missense	probably damaging	0
RBM33	chr7	E424A	missense	possible damaging	0
RRP1B	chr21	I169V	missense	benign	0
SARDH	chr9	P804L	missense	benign	0
SH2D3A	chr19	T178M	missense	benign	0
SLC6A17	chr1	I142T	missense	probably damaging	0
SLC9A5	chr16	R270C	missense	probably damaging	0
TJP3	chr19	R162H	missense	probably damaging	0
TRPM4	chr19	G238S	missense	probably damaging	0
VPS11	chr11	E714K	missense	benign	0
ZFP64	chr20	S399R	missense	probably damaging	0

Table 12. Novel variants in DNMs containing genes identified in a cohort of 220 FALS

Gene	Chr	aa change	mutation	polyPhen
ASCC3	chr6	p.N135K	missense	benign
C10orf90	chr10	p.R220P	missense	probably damaging
CLCN2	chr3	p.I146T	missense	benign
CLCN2	chr3	p.P738L	missense	benign
CLCN2	chr3	p.R66W	missense	possibly damaging
EPHA2	chr1	p.A536V	missense	probably damaging
EPHA2	chr1	p.R159C	missense	probably damaging
EPHA2	chr1	p.R566H	missense	benign
EPHA2	chr1	p.S169P	missense	benign
FGD5	chr3	p.I629V	missense	benign
FGD5	chr3	p.P1325R	missense	benign
LTV1	chr6	p.D65N	missense	probably damaging
MFSD12	chr19	p.D475Y	missense	possibly damaging
MFSD12	chr19	p.G452D	missense	probably damaging
MFSD12	chr19	p.R218W	missense	probably damaging
MFSD12	chr19	p.S386L	missense	probably damaging
ODZ3	chr4	p.G1845R	missense	possibly damaging
ODZ3	chr4	p.I594V	missense	benign
ODZ3	chr4	p.M692I	missense	benign
PAPPA	chr9	p.H700R	missense	benign
PAPPA	chr9	p.N514S	missense	probably damaging
RBM33	chr7	p.Q342L	missense	probably damaging
RRP1B	chr21	p.E218K	missense	probably damaging
SARDH	chr9	p.V498M	missense	benign
SH2D3A	chr19	p.V416I	missense	possibly damaging
SLC6A17	chr1	p.G716S	missense	probably damaging
SLC6A17	chr1	p.V243I	missense	benign
TJP3	chr19	p.P52R	missense	benign
TRPM4	chr19	p.A863V	missense	benign
ZFP64	chr20	p.S578T	missense	probably damaging

Figure 17. Biological network of the direct and indirect interactions among gene products of our ALS trios candidate genes Colours indicate significance of participation in the protein-protein interaction network.

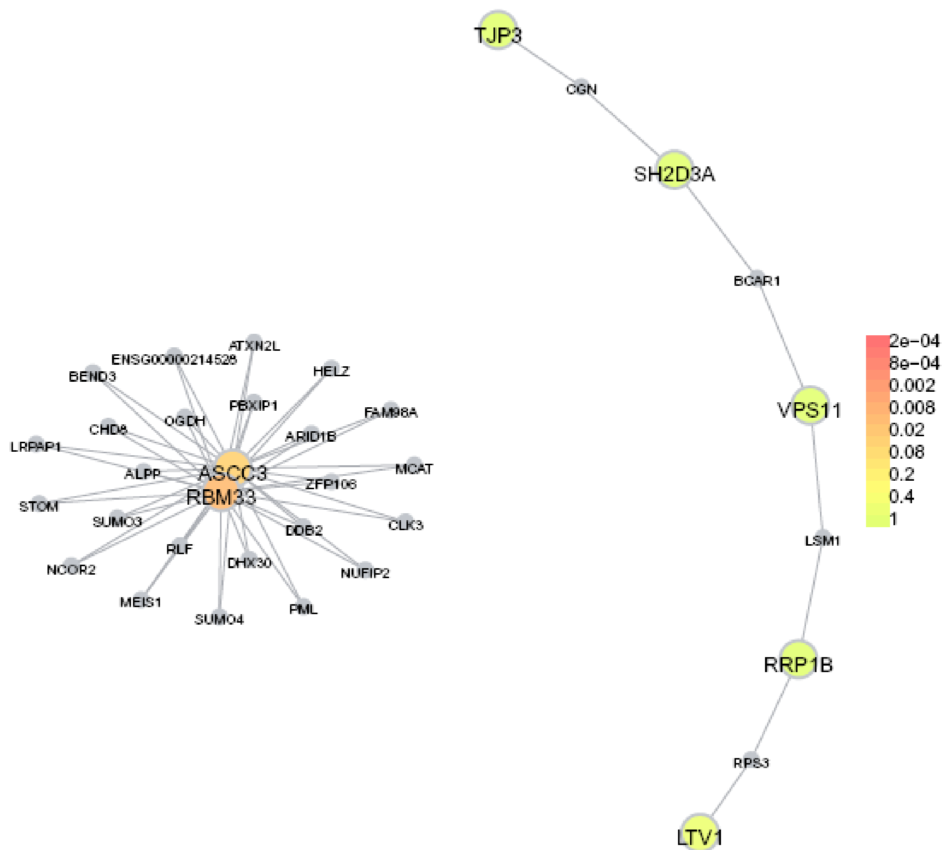
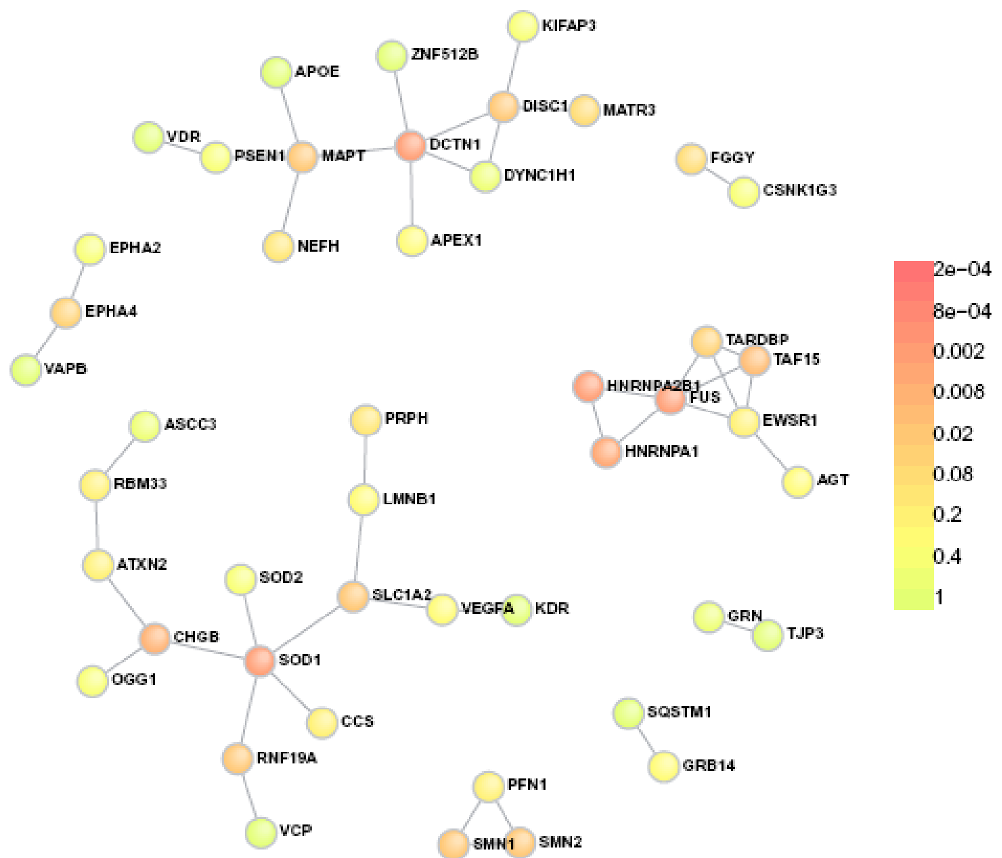


Figure 18. Biological network of the direct interactions among proteins from our candidate genes and known ALS-related genes created using using DAPPLE. Colours indicate significance of participation in the protein-protein interaction network.



5. DISCUSSION

The recent success achieved by exome-sequencing in the identification of many causative genes in Mendelian disorders led to the application of this approach also to neurological complex diseases in which the only use of conventional approaches has failed. The exome-sequencing approach has already been applied in ALS with the identification of different novel causative genes, including *VCP, hnRNPA2/B1* and *MATR3* [76,77,89].

Our results confirm that WES, combined with different strategies for study design and analysis, is an effective approach for the identification of ALS-causing genes.

By performing WES on two large ALS families displaying a dominant pattern of inheritance, we identified causative mutations in *PFN1* gene, which encodes for a small actin-binding protein. The reported success in disease-gene identification in these two families was accomplished by the use of a well-designed strategy: a) exome sequencing was performed on the two affected members of each family with maximum genetic distance, to select only the shared variants and to reduce the number of private variants to retain; b) parallel use of linkage analysis increased the efficiency of variant filtering by excluding variants lying outside of linkage peaks.

The subsequent screening of *PFN1* gene in a large cohort of FALS index cases resulted in the detection of additional mutations and assessed the mutational frequency of *PFN1* to 2,6% in familial cases, a value comparable to that reported for other ALS causative genes such as *VCP.* The analysis extended to sporadic cases identified only the p.E117G variant in 3/1984 SALS.

The pathogenicity of *PFN1* variants in ALS was supported by three different genetic evidences: 1) co-segregation was assessed in three families (#1, #2, #3) and indicated for PFN1 mutations a high degree of penetrance; 2) evaluation of the evolutionary sequence conservation in conjunction with

different bioinformatic tools predicted potential damaging effects in terms of biological function; 3) three out of the 4 variants were absent in a large control population (more than 9000 individuals), while one variant (p.E117G) was detected in controls at a significant lower frequency.
In addition to genetic data, several evidences from functional studies supported the predicted damaging impact of PFN1 variants. PFN1 mutants formed ubiquitin-positive inclusions in transfected N2A and primary motor neurons, analogous to those observed for other proteins implicated in ALS, and showed several functional differences compared to the wild-type protein. In particular, a reduction of actin binding ability and an inhibition of axonal outgrowth were observed. Notably, in line with the genetic observations, the variant p.E117G exhibited only mild functional defects suggesting that is probably a less pathogenic mutant.
The discovery of *PFN1* further proves that the combination of traditional genetic approaches, such as linkage analysis, with exome sequencing proves is highly successful and powerful for detecting novel genes associated with familial ALS.

Despite these advances, a wider application of this strategy for the identification of causative mutations in ALS is still difficult. Informative pedigrees, with multiple affected members available for testing, are rare for a neurodegenerative disease like ALS, characterized by late-onset and short survival time. Therefore, the adoption of novel and different analytical strategies that increase the probability of identifying causative rare variants is required to study successfully also unrelated index familial cases, which instead can be collected in a larger number. After performing exome-sequencing in a large cohort of 363 index FALS, we applied for the first time a novel gene-based case-control analysis to identify genes showing an excess of rare and predicted damaging variants in cases compared to controls. This strategy was designed to combine information across

multiple variation sites within a gene, taking into account the genetic heterogeneity of ALS. The approach was successful as demonstrated by the identification of *TUBA4A*, encoding for a member of the alpha-tubulin family, as another candidate ALS gene. The disease-related variants affected strongly conserved residues and most variants were absent in over 13,000 controls. Moreover, the functional characterization of mutants provided strong evidence of deleterious effects on microtubule dynamics: a) inefficient formation of alpha/beta tubulin dimers *in vitro*; b) decreased incorporation into microtubules in cultured cells. In particular, the mutant p.W407X protein appeared totally deficient in forming heterodimers or in incorporating into the microtubule network, because of the lack of the C-terminal region containing the binding site for beta-tubulin. Moreover, this mutant showed aggregation propensities analogous to those of other proteins implicated in ALS, suggesting that this mutations could act by a different pathological mechanism, such as sequestration of tubulin-binding proteins into the aggregates or impairment of ubiquitin-proteasome system. Similarly to what observed for the p.E117G variant in *PFN1*, the mutant TUBA4A p.G43V, found in a sporadic ALS case but also in a control individual, did not display significant functional differences from the wild-type protein.

Finally, the third strategy of WES that we applied involved a family-based trios design in order to identify DNMs in 32 sporadic ALS patients. In support of our working hyphotesis, previous studies had provided evidence for a significant involvement of DNMs in complex neurological diseases such as schizophrenia, autism, and sporadic mental retardation and more recently in ALS. By sequencing the exome of the patient as well as his/her parents, we selected *de novo* candidates by filtering out all inherited variants. The advantage of this trios-approach is that it yields a limited number of potential pathogenic variants, because multiple random *de novo*

events occurring within an individual are extremely unlikely and each exome contains an average of only 0–3 DNMs. In line with the reported DNMs ratio, we did not identify more than 2/3 DNMs in 16/32 trios, while we failed to detect DNMs in the remaining 16 samples. Bioinformatical analysis of the total 25 genes containing DNMs showed potential inter-connections and an enrichment of specific biological functions, in particular the transport and the GTPase regulatory activity.

In conclusion, our discoveries further support the complexity of ALS genetics, characterized by a great heterogeneity, with many different genes causing the same disease phenotype. Interestingly both *PFN1* and *TUBA4A*, identified by exome-sequencing approach in familial cases, are implicated in the assembly and maintenance of the cytoskeletal network. Profilin-1 is a ubiquitous small actin-binding protein, which catalyzes the ADP/ATP exchange of monomeric actin (G-actin) and the addition of ATP-bound G-actin at the barbed (+) end of actin filaments resulting in actin polymerization [87,88]. The fine regulation of actin dynamics, modulated by many other G- and F-actin-binding proteins (ABPs), is fundamental in a variety of cellular process and a malfunction of cytoskeletal components causes several human diseases, including neurological disorders. Particularly, in neuronal cells the rapid assembly and disassembly of F-actin is particularly crucial in dendritic spine for synaptic plasticity and in the growth cone for the axon outgrowth and guidance to specific targets during neuronal development. By demonstrating that ALS-related PFN1 mutants alter actin dynamics and inhibit axon outgrowth, we demonstrated a possible mechanism by which *PFN1* may contribute to ALS pathogenesis. A similar inhibition on neurite outgrowth was observed also for mutant SOD1[91] and TDP-43[92], while an alteration of spine density and morphology was detected in primary neurons from FUS deficient mice [93].

These observations suggest a common pathogenic feature among different ALS genes. The second gene identified by WES encodes for Tubulin alpha 4a, which is a major component of microtubules (MT). MTs are implicated in a large spectrum of cellular processes, from mytosis and cellular motility to intracellular transport and secretion. In neurons microtubule cytoskeleton supports the remodeling of dendritic spines in response to synaptic transmission, the neurite branching, the growth cone advance for axonal elongation and guidance, and the intracellular trafficking to the pre- and post-synaptic compartments [94]. Alterations of the axonal microtubule network seems to impair the MT-based anterograde and retrograde transport contributing finally to axonal swelling. The axonal transport is particular crucial in motor neurons, which have long axons that can extend for up to a meter to reach their target, and therefore any disturbance of axonal transport may have severe consequences on the function and survival of motor neurons. We demonstrated severe functional defects for TUBA4A mutants, including impaired heterodimerization and defective incorporation into microtubules, which could trigger motor neuron degeneration in ALS. Interestingly, mutations in other tubulin genes (TUBA1A, TUBA8, TUBB2B, TUBB3, TUBG1,TUBB4A) have been linked to several neurodevelopmental and neurodenegerative diseases [95]. Most mutant tubulin proteins show functional defects similar to the ones observed for TUBA4A mutants, such as impaired dimerization and microtubule incorporation, and yet cause very different phenotypes. The difference could be explained by a differential temporal expression of tubulin during development: most α- and β-tubulin genes responsible for developmental disorders are highly expressed early during brain development and decrease with age, whereas TUBA4A levels increase dramatically (>50-fold) with age. Furthermore, the progressive motor neuronopathy (pmn) mutant mouse, a commonly used model for human motor neuron disease, is the result of a mutation in the Tubulin-specific

Chaperone E (TBCE) gene [96]. A downregulation of genes encoding for α-tubulin subunits has been reported in spinal motor neurons of SALS patients [97]

The identification of *PFN1* and *TUBA4A* as novel causative genes for ALS reinforces the hypothesis that cytoskeletal defects have a major role in disease pathogenesis. Other genes encoding for cytoskeletal proteins, *DNCT1, PRPH,* and *NEFH*, have been associated to ALS. From a neuropathological point of view spheroids inclusions containing neurofilaments and kinesins are found in affected motor neurons. Although these alterations were considered for a long time a secondary consequence from the neuronal degeneration, recent studies on animal models have demonstrated that alterations of intermediate filaments can trigger cell death in neurons.

6. CONCLUSIONS

The identification of novel disease genes represents the first step to a better understanding of the pathological mechanisms and pathways underlying the disease, which in turn serves as a starting point for developing therapeutic interventions. Exome sequencing, with its unbiased-nature, has significantly progressed and transformed ALS genetics in the last few years. By applying tailored study designs and strategies for variant prioritization in exome-sequencing, we were able to identify two novel causative genes, *PFN1* and *TUBA4A*, both affecting the cytoskeletal network.

In particular, exome sequencing combined with linkage analysis, has proven to be the highly successful approach, as demonstrated by the identification of *PFN1* in our study and by the increasing number of ALS-causing genes identified in the last few years.

The second gene, *TUBA4A*, was identified by using a novel exome-wide rare variant analysis in a large sample cohort of unrelated FALS, not amendable to conventional approaches. The analysis, which allows to identify genes showing a significant excess of rare damaging variants in cases compared to controls, has demonstrated to be effective in ALS. Although *TUBA4A* was the top hit, the study identified other candidate genes, which will be further investigated by performing additional mutational screenings in ALS cohort and functional studies. Other genetic analyses are needed also for *TUBA4A* to assess its potential contribution to the genetics of FTD, in view of the ALS-FTD phenotype observed in 3 out of 6 mutated patients.

In conclusion, exome sequencing has proven to be a robust method to study a disorder with great genetic and phenotypic heterogeneity like familial ALS. Although these advancements, it's important to underline that genes identified by WES so far (*PFN1, TUBA4A, VCP, MATR3,*

hnRNPA2/B1) account only for a very small percentage of FALS cases, while they don't seem to contribute significantly to the aetiology of sporadic forms. One future challenge will be to systematically study the role of variations in the non-coding regions by full genome sequencing. As demonstrated by C9orf72, it's possible that causative mutations for the remaining genetically unknown FALS may localize in non coding regions which are not captured in WES.

In comparison to FALS, the exome sequencing studies on sporadic ALS are few, although increasing evidence points to a role of rare variants for explaining a large part of SALS heritability. To improve the chances of successfully discovering rare variants associated with ALS susceptibility, we explored a trio-based approach aimed at the identification of *de novo* coding mutations. Similarly to a study recently published [78], our results confirm that recurrently mutated genes among different trios are rare for a genetically heterogeneous disease like ALS. Additional *in silico* evaluations of protein-protein interactions and pathway-based analyses can provide supportive evidences useful for candidate genes prioritization. We are confident that increasing the number of ALS trios analyzed, although difficult for the limited availability of parental samples in an adult-onset disease, and integrating data from multiple similar studies will help to improve the disease gene discovery efforts in sporadic ALS.

7. REFERENCES

1. Shaw CE, al-Chalabi A, Leigh N. Progress in the pathogenesis of amyotrophic lateral sclerosis. *Curr Neurol Neurosci Rep.* 2001;1:69–76.

2. Meininger V. ALS, what new 144 years after Charcot? *Arch Ital Biol.* 2011;149:29-37.

3. Kumar DR, Aslinia F., Yale SH, Mazza JJ. Jean-Martin Charcot: The Father of Neurology. *Clin Med Res.* 2011; 9:46-49.

4. Goetz CG. Amyotrophic lateral sclerosis: early contributions of Jean-Martin Charcot. *Muscle Nerve* 2000;23:336-343.

5. Gordon PH. Amyotrophic Lateral Sclerosis: An update for 2013 Clinical Features, Pathophysiology, Management and Therapeutic Trials. *Aging Dis*. 2013; 4:295-310.

6. Haverkamp LJ, Appel V, Appel SH. Natural history of amyotrophic lateral sclerosis in a database population. Validation of a scoring system and a model for survival prediction. *Brain* 1995;118:707-719.

7. Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. *N Engl J Med*. 2001; 344: 1688-1700.

8. Silani V, Messina S, Poletti B, Morelli C, Doretti A, Ticozzi N, Maderna L. The diagnosis of Amyotrophic lateral sclerosis in 2010. *Arch Ital Biol.* 2011; 149:5-27.

9. Wijesekera LC, Mathers S, Talman P, Galtrey C, Parkinson MH, Ganesalingam J, Willey E, Ampong MA, Ellis CM, Shaw CE, Al-Chalabi A, Leigh PN. Natural history and clinical features of the flail arm and flail leg ALS variants. *Neurology* 2009; 72:1087-1094.

10. Gordon PH, Cheng B, Katz IB, Pinto M, Hays AP, Mitsumoto H, Rowland LP. The natural history of primary lateral sclerosis. *Neurology* 2006;66: 647-653.

11. Mannen T, Iwata M, Toyokura Y, Nagashima K. Preservation of a certain motoneurone group of the sacral cord in amyotrophic lateral sclerosis: its clinical significance. *J Neurol Neurosurg Psychiatry* 1977; 40:464–469.

12. Brockington A, Ning K, Heath PR, Wood E, Kirby J, Fusi N, Lawrence N, Wharton SB, Ince PG, Shaw PJ. Unravelling the enigma of selective vulnerability in neurodegeneration: motor neurons resistant to degeneration in ALS show distinct gene expression characteristics and decreased susceptibility to excitotoxicity. *Acta Neuropathol*. 2013;125:95-109.

13. Phukan J, Elamin M, Bede P, Jordan N, Gallagher L, Byrne S, Lynch C, Pender N, Hardiman O. The syndrome of cognitive impairment in amyotrophic lateral sclerosis: a population-based study. *J.Neurol. Neurosurg. Psychiatry* 2012;83:102-108.

14. Ferrari R, Kapogiannis D, Huey ED, Momeni P. FTD and ALS: a tale of two diseases. *Curr Alzheimer Res*. 2011;8(3):273-294.

15. Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* 2005;65:586-590.

16. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology.* 1998;51: 1546-54.

17. Ling SC, Polymenidou M, Cleveland DW. Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. *Neuron* 2013;79:416-438.

18. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, Chou TT, Bruce J, Schuck T, Grossman M, Clark CM, McCluskey LF, Miller BL, Masliah E, Mackenzie IR, Feldman H, Feiden W, Kretzschmar HA, Trojanowski JQ, Lee VM. Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 2006;314:130-133.

19. Munoz DG, Neumann M, Kusaka H, Yokota O, Ishihara K, Terada S, Kuroda S, Mackenzie IR. FUS pathology in basophilic inclusion body disease. *Acta Neuropathol.*2009;118:617-627.

20. DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A, Sengdy P, Hsiung GY, Karydas A, Seeley WW, Josephs KA, Coppola G, Geschwind DH, Wszolek ZK, Feldman H, Knopman DS, Petersen RC, Miller BL, Dickson DW, Boylan KB, Graff-Radford NR,

Rademakers R. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011;72:245-256.

21. Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita VM, Kaivorinne AL, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M; ITALSGEN Consortium, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ et al.. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011;72:257-268.

22. Miller RG, Munsat TL, Swash M, Brooks BR. Consensus guidelines for the design and implementation of clinical trials in ALS. World Federation of Neurology committee on Research. *J Neurol Sci.*1999;169: 2-12.

23. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J Neurol Sci.* 1994;12Suppl:96-107.

24. Brooks BR, Miller RG, Swash M, Munsat TL; World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord.* 2000;1:293-299.

25. Abrahams S, Newton J, Niven E, Foley J, Bak TH. Screening for cognition and behaviour changes in ALS. *Amyotroph Lateral Scler Frontotemporal Degener.* 2014;15:9-14.

26. Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, Rijsdijk F. An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry* 2010;81:1324-1236.

27. Hanby MF, Scott KM, Scotton W, Wijesekera L, Mole T, Ellis CE, Leigh PN, Shaw CE, Al-Chalabi A. The risk to relatives of patients with sporadic amyotrophic lateral sclerosis. *Brain* 2011;134:3454-3457.

28. Al-Chalabi A, Hardiman O. The epidemiology of ALS: a conspiracy of genes, environment and time. *Nat Rev Neurol.* 2013;9:617-628.

29. Byrne S, Walsh C, Lynch C, Bede P, Elamin M, Kenna K, McLaughlin R, Hardiman O. Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2011;82:623-627.

30. Majoor-Krakauer D, Willems PJ, Hofman A. Genetic epidemiology of amyotrophic lateral sclerosis. *Clin Genet.*2013; 63: 83-101.

31. Byrne S, Elamin M, Bede P, Hardiman O. Absence of consensus in diagnostic criteria for familial neurodegenerative diseases. *J Neurol Neurosurg Psychiatry* 2012;83:365-367.

32. Mulder DW, Kurland LT, Offord KP, Beard CM. Familial adult motor neuron disease: amyotrophic lateral sclerosis. *Neurology* 1986;36:511-517.

33. Rothstein JD. Current hypotheses for the underlying biology of amyotrophic lateral sclerosis. *Ann Neurol.* 2009;65:S3-S9.

34. Kuzuhara S, Kokubo Y, Sasaki R, Narita Y, Yabana T, Hasegawa M, Iwatsubo T. Familial amyotrophic lateral sclerosis and parkinsonism-dementia complex of the Kii Peninsula of Japan: clinical and neuropathological study and tau analysis. *Ann Neurol*. 2011;49:501-511.

35. Spencer PS, Hugon J, Ludolph A, Nunn PB, Ross SM, Roy DN, Schaumburg HH. Discovery and partial characterization of primate motor-system toxins. *Ciba Found Symp*. 1987;126:221-38.

36. Yoshida S, Uebayashi Y, Kihira T, Kohmoto J, Wakayama I, Taguchi S, Yase Y. Epidemiology of motor neuron disease in the Kii Peninsula of Japan, 1989-1993: active or disappearing focus? *J Neurol Sci.* 1998;155: 146-155.

37. Caller TA, Doolin JW, Haney JF, Murby AJ, West KG, Farrar HE, Ball A, Harris BT, Stommel EW. A cluster of amyotrophic lateral sclerosis in New Hampshire: a possible role for toxic cyanobacteria blooms. *Amyotroph Lateral Scler.* 2009;Suppl 2:101-108.

38. Masseret E, Banack S, Boumédiène F, Abadie E, Brient L, Pernet F, Juntas-Morales R, Pageot N, Metcalf J, Cox P, Camu W; French Network on ALS Clusters Detection and Investigation. Dietary BMAA exposure in an amyotrophic lateral sclerosis cluster from southern France. *PLoS One* 2013;8(12):e83406.

39. Johnson FO, Atchison WD. The role of environmental mercury, lead and pesticide exposure in development of amyotrophic lateral sclerosis. *Neurotoxicology* 2009;30:761-765.

40. Trojsi F, Monsurrò MR, Tedeschi G. Exposure to environmental toxicants and pathogenesis of amyotrophic lateral sclerosis: state of the art and research perspectives. *Int J Mol Sci*. 2013;14:15286-15311.

41. Callaghan B, Feldman D, Gruis K, Feldman E. The association of exposure to lead, mercury, and selenium and the development of amyotrophic lateral sclerosis and the epigenetic implications. *Neurodegener Dis*. 2011;8:1-8.

42. Beghi E. Are professional soccer players at higher risk for ALS? *Amyotroph Lateral Scler Frontotemporal Degener.* 2013;14:501-516.

43. Alfahad T, Nath A. Retroviruses and amyotrophic lateral sclerosis. *Antiviral Res.* 2011;99:180-187.

44. Figueroa-Romero C, Hur J, Bender DE, Delaney CE, Cataldo MD, Smith AL, Yung R, Ruden DM, Callaghan BC, Feldman EL. Identification of epigenetically altered genes in sporadic amyotrophic lateral sclerosis. *PLoS One* 202;7:e52672.

45. Leblond CS, Kaneb HM, Dion PA, Rouleau GA. Dissection of genetic factors associated with amyotrophic lateral sclerosis. *Exp Neurol.* 2014;262:91-101.

46. Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX, et al. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 1993;362: 59-62.

47. Ticozzi N, Tiloca C, Morelli C, Colombrita C, Poletti B, Doretti A, Maderna L, Messina S, Ratti A, Silani V. Genetics of familial Amyotrophic lateral sclerosis. *Arch Ital Biol.* 2011;149:65-82.

48. Rademakers R. C9orf72 repeat expansions in patients with ALS and FTD. *Lancet Neurol.* 2012;11:297-298.

49. Konno T, Shiga A, Tsujino A, Sugai A, Kato T, Kanai K, Yokoseki A, Eguchi H, Kuwabara S, Nishizawa M, Takahashi H, Onodera O. Japanese amyotrophic lateral sclerosis patients with GGGGCC hexanucleotide repeat expansion in C9ORF72. *J Neurol Neurosurg Psychiatry* 2013;84:398-401.

50. Alavi A, Nafissi S, Rohani M, Shahidi G, Zamani B, Shamshiri H, Safari I, Elahi E. Repeat expansion in C9ORF72 is not a major cause of amyotrophic lateral sclerosis among Iranian patients. *Neurobiol Aging* 2014;35:267.e1-e7.

51. Cruts M, Gijselinck I, Van Langenhove T, van der Zee J, Van Broeckhoven C. Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum. *Trends Neurosci*. 2013;36:450-459.

52. Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RA, Levina V, Halloran MA, Gleeson PA, Blair IP, Soo KY, King AE, Atkin JD. C9ORF72, implicated in amytrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet.* 2014;23:3579-3595.

53. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, Vande Velde C, Bouchard JP, Lacomblez L, Pochigaeva K, Salachas F, Pradat PF, Camu W, Meininger V, Dupre N, Rouleau GA. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet.* 2008;40:572-574.

54. Colombrita C, Onesto E, Tiloca C, Ticozzi N, Silani V, Ratti A. RNA-binding proteins and RNA metabolism: a new scenario in the pathogenesis of Amyotrophic lateral sclerosis. *Arch Ital Biol.* 2011;149:83-99.

55. Lattante S, Rouleau GA, Kabashi E. TARDBP and FUS Mutations Associated with Amyotrophic Lateral Sclerosis: Summary and Update. *Hum Mutat.* 2013;34:812-826.

56. Chiò A, Borghero G, Pugliatti M, Ticca A, Calvo A, Moglia C, Mutani R, Brunetti M, Ossola I, Marrosu MG, Murru MR, Floris G, Cannas A, Parish LD, Cossu P, Abramzon Y, Johnson JO, Nalls MA, Arepalli S, Chong S, Hernandez DG, Traynor BJ, Restagno G; Italian Amyotrophic Lateral Sclerosis Genetic (ITALSGEN) Consortium. Large proportion of amyotrophic lateral sclerosis cases in Sardinia due to a single founder mutation of the TARDBP gene. *Arch Neurol.* 2011;68:594-598.

57. Abalkhail H, Mitchell J, Habgood J, Orrell R, de Belleroche J. A new familial amyotrophic lateral sclerosis locus on chromosome 16q12.1-16q12.2. *Am J Hum Genet.* 2003;73:383-389.

58. Kwiatkowski TJ, Jr., Bosco DA, Leclerc AL, Tamrazian E, Vanderburg CR, Russ C, Davis A, Gilchrist J, Kasarskis EJ, Munsat T, Valdmanis P, Rouleau GA, Hosler BA, Cortelli P, de Jong PJ, Yoshinaga Y, Haines JL, Pericak-Vance MA, Yan J, Ticozzi N, Siddique T, McKenna-Yasek D, Sapp PC, Horvitz HR, Landers JE, Brown RH, Jr. Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* 2009;323:1205-1208.

59. Vance C, Rogelj B, Hortobagyi T, De Vos KJ, Nishimura AL, Sreedharan J, Hu X, Smith B, Ruddy D, Wright P, Ganesalingam J, Williams KL, Tripathi V, Al-Saraj S, Al-Chalabi A, Leigh PN, Blair IP, Nicholson G, de Belleroche J, Gallo JM, Miller CC, Shaw CE. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* 2009;323:1208-1211.

60. Ferrara N. Vascular endothelial growth factor and the regulation of angiogenesis. *Recent Prog Horm Res*. 2000;55:15-35.

61. Lambrechts D, Storkebaum E, Morimoto M, Del-Favero J, Desmet F, Marklund SL, Wyns S, Thijs V, Andersson J, van Marion I, Al-Chalabi A, Bornes S, Musson R, Hansen V, Beckman L, Adolfsson R, Pall HS, Prats H, Vermeire S, Rutgeerts P, Katayama S, Awata T, Leigh N, Lang-Lazdunski L, Dewerchin M, Shaw C, Moons L, Vlietinck R, Morrison KE, Robberecht W, Van Broeckhoven C, Collen D, Andersen PM, Carmeliet P.

VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nat Genet*. 2003;34: 383-394.

62. Storkebaum E, Lambrechts D, Dewerchin M, Moreno-Murciano MP, Appelmans S, Oh H, Van Damme P, Rutten B, Man WY, De Mol M, Wyns S, Manka D, Vermeulen K, Van Den Bosch L, Mertens N, Schmitz C, Robberecht W, Conway EM, Collen D, Moons L, Carmeliet P. Treatment of motoneuron degeneration by intracerebroventricular delivery of VEGF in a rat model of ALS. *Nat Neurosci*. 2005;8:85-92.

63. McLaughlin RL, Phukan J, McCormack W, Lynch DS, Greenway M, Cronin S, Saunders J, Slowik A, Tomik B, Andersen PM, Bradley DG, Jakeman P, Hardiman O. Angiogenin levels and ANG genotypes: dysregulation in amyotrophic lateral sclerosis. *PLoS One*. 2010;5:e15402.

64. Greenberg DA, Stewart WC, Rowland LP. Paraoxonase genes and susceptibility to ALS. *Neurology.* 2009;73:11-12.

65. Ticozzi N, LeClerc AL, Keagle PJ, Glass JD, Wills AM, van Blitterswijk M, Bosco DA, Rodriguez-Leyva I, Gellera C, Ratti A, Taroni F, McKenna-Yasek D, Sapp PC, Silani V, Furlong CE, Brown RH Jr, Landers JE. *Ann Neurol*. 2010;68:102-107.

66. Goodall EF, Greenway MJ, van Marion I, Carroll CB, Hardiman O, Morrison KE. Association of the H63D polymorphism in the hemochromatosis gene with sporadic ALS. *Neurology*. 2005;65:934-937.

67. Wang MD, Gomes J, Cashman NR, Little J, Krewski D. Intermediate CAG repeat expansion in the ATXN2 gene is a unique genetic risk factor for ALS--a systematic review and meta-analysis of observational studies. *PLoS One* 2014; 9:e105534.

68. Siddique T, Figlewicz DA, Pericak-Vance MA, Haines JL, Rouleau G, Jeffers AJ, Sapp P, Hung WY, Bebout J, McKenna-Yasek D, Deng G, Horvitz RH, Gusella JF, Brown RH, Roses AD and Collaborators. Linkage of a gene causing familial amyotrophic lateral sclerosis to chromosome 21 and evidence of genetic-locus heterogeneity. *N Engl J Med*. 1991;324: 1381-1384.

69. Xiao S, McLean J, Robertson J. Neuronal intermediate filaments and ALS: a new look at an old question. *Biochim Biophys Acta* 2006; 1762:11-12.

70. Ticozzi N, Vance C, Leclerc AL, Keagle P, Glass JD, McKenna-Yasek D, Sapp PC, Silani V, Bosco DA, Shaw CE, Brown RH Jr, Landers JE. Mutational analysis reveals the FUS homolog TAF15 as a candidate gene for familial amyotrophic lateral sclerosis. *Am J Med Genet B Neuropsychiatr Genet*. 2011;156:285-290.

71. Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, Johnson L, Veldink JH, van Es MA, van den Berg LH, Robberecht W, Van Damme P, Hardiman O, Farmer AE, Lewis CM, Butler AW, Abel O, Andersen PM, Fogh I, Silani V, Chiò A, Traynor BJ, Melki J, Meininger V, Landers JE, McGuffin P, Glass JD, Pall H, Leigh PN, Hardy J, Brown RH Jr, Powell JF, Orrell RW, Morrison KE, Shaw PJ, Shaw CE, Al-Chalabi A. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol.* 2010;9:986-994.

72. Vance C, Al-Chalabi A, Ruddy D, Smith BN, Hu X, Sreedharan J, Siddique T, Schelhaas HJ, Kusters B, Troost D, et al. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain* 2006;129:868–876.

73. Fogh I, Ratti A, Gellera C, Lin K, Tiloca C, Moskvina V, Corrado L, Sorarù G, Cereda C, Corti S, Gentilini D, Calini D, Castellotti B, Mazzini L, Querin G, Gagliardi S, Del Bo R, Conforti FL, Siciliano G, Inghilleri M, Saccà F, Bongioanni P, Penco S, Corbo M, Sorbi S, Filosto M, Ferlini A, Di Blasio AM, Signorini S, Shatunov A, Jones A, Shaw PJ, Morrison KE, Farmer AE, Van Damme P, Robberecht W, Chiò A, Traynor BJ, Sendtner M, Melki J, Meininger V, Hardiman O, Andersen PM, Leigh NP, Glass JD, Overste D, Diekstra FP, Veldink JH, van Es MA, Shaw CE, Weale ME, Lewis CM, Williams J, Brown RH, Landers JE, Ticozzi N, Ceroni M, Pegoraro E, Comi GP, D'Alfonso S, van den Berg LH, Taroni F, Al-Chalabi A, Powell J, Silani V; SLAGEN Consortium and Collaborators. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet*. 2014;23:2220-2231.

74. Landers JE, Melki J, Meininger V, Glass JD, van den Berg LH, van Es MA, Sapp PC, van Vught PW, McKenna-Yasek DM, Blauw HM, Cho TJ, Polak M, Shi L, Wills AM, Broom WJ, Ticozzi N, Silani V, Ozoguz A, Rodriguez-Leyva I, Veldink JH, Ivinson AJ, Saris CG, Hosler BA, Barnes-Nessa A, Couture N, Wokke JH, Kwiatkowski TJ Jr, Ophoff RA, Cronin S, Hardiman O, Diekstra FP, Leigh PN, Shaw CE, Simpson CL, Hansen VK, Powell JF, Corcia P, Salachas F, Heath S, Galan P, Georges F, Horvitz HR, Lathrop M, Purcell S, Al-Chalabi A, Brown RH Jr. Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A*. 2009;106:9004-9009.

75. Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet*. 2011;129:351-370.

76. Johnson JO, Mandrioli J, Benatar M, Abramzon Y, Van Deerlin VM, Trojanowski JQ, Gibbs JR, Brunetti M, Gronka S, Wuu J, Ding J, McCluskey L, Martinez-Lage M, Falcone D, Hernandez DG, Arepalli S, Chong S, Schymick JC, Rothstein J, Landi F, Wang YD, Calvo A, Mora G, Sabatelli M, Monsurrò MR, Battistini S, Salvi F, Spataro R, Sola P, Borghero G; ITALSGEN Consortium, Galassi G, Scholz SW, Taylor JP, Restagno G, Chiò A, Traynor BJ. Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron*. 2010;68:857-864.

77. Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Molliex A, Kanagaraj AP, Carter R, Boylan KB, Wojtas AM, Rademakers R, Pinkus JL, Greenberg SA, Trojanowski JQ, Traynor BJ, Smith BN, Topp S, Gkazi AS, Miller J, Shaw CE, Kottlors M, Kirschner J, Pestronk A, Li YR, Ford AF, Gitler AD, Benatar M, King OD, Kimonis VE, Ross ED, Weihl CC, Shorter J, Taylor JP. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* 2013;495:467-473.

78. Chesi A, Staahl BT, Jovičić A, Couthouis J, Fasolino M, Raphael AR, Yamazaki T, Elias L, Polak M, Kelly C, Williams KL, Fifita JA, Maragakis NJ, Nicholson GA, King OD, Reed R, Crabtree GR, Blair IP, Glass JD, Gitler AD. Exome sequencing to identify de novo mutations in sporadic ALS trios. *Nat Neurosci.* 2013;16:851-855.

79. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci.* 2014;17:17-23.

80. Hewitt C, Kirby J, Highley JR, Hartley JA, Hibberd R, Hollinger HC, Williams TL, Ince PG, McDermott CJ, Shaw PJ. Novel FUS/TLS mutations and pathology in familial and sporadic amyotrophic lateral sclerosis. *Arch Neurol*. 2010;67:455-461.

81. Gitler AD, Shorter J. RNA-binding proteins with prion-like domains in ALS and FTLD-U. *Prion*. 2011;5:179-187.

82. Shaw BF, Valentine JS. How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? *Trends Biochem Sci*. 2007;32:78-85.

83. Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, Dejesus-Hernandez M, van Blitterswijk MM, Jansen-West K, Paul JW 3rd, Rademakers R, Boylan KB, Dickson DW, Petrucelli L. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 2013;77:639-646.

84. Robberecht W, Philips T. The changing scene of amyotrophic lateral sclerosis. *Nat Rev Neurosci.* 2013;14:248-264.

85. Julien JP, Beaulieu JM. Cytoskeletal abnormalities in amyotrophic lateral sclerosis: beneficial or detrimental effects? *J Neurol Sci*. 2000;180:7-14.

86. Fallini C, Bassell GJ, Rossoll W. High-efficiency transfection of cultured primary motor neurons to study protein localization, trafficking, and function. *Mol Neurodegener*. 2010;5:17.

87. Mockrin SC, Korn ED. Acanthamoeba profilin interacts with G-actin to increase the rate of exchange of actin-bound adenosine 5′-triphosphate. *Biochemistry 1980;*19:5359–5362.

88. Suetsugu S, Miki H, Takenawa T. The essential role of profilin in the assembly of actin for microspike formation. *EMBO J*. 1988;17:6516–6526.

89. Johnson JO, Pioro EP, Boehringer A, Chia R, Feit H, Renton AE, Pliner HA, Abramzon Y, Marangi G, Winborn BJ, Gibbs JR, Nalls MA, Morgan S, Shoai M, Hardy J, Pittman A, Orrell RW, Malaspina A, Sidle KC, Fratta P, Harms MB, Baloh RH, Pestronk A, Weihl CC, Rogaeva E, Zinman L, Drory VE, Borghero G, Mora G, Calvo A, Rothstein JD; ITALSGEN Consortium, Drepper C, Sendtner M, Singleton AB, Taylor JP, Cookson MR, Restagno G, Sabatelli M, Bowser R, Chiò A, Traynor BJ. Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat Neurosci.* 2014;17:664-666.

90. Van Hoecke A, Schoonaert L, Lemmens R, Timmers M, Staats KA, Laird AS, Peeters E, Philips T, Goris A, Dubois B, Andersen PM, Al-Chalabi A, Thijs V, Turnley AM, van Vught PW, Veldink JH, Hardiman O, Van Den Bosch L, Gonzalez-Perez P, Van Damme P, Brown RH Jr, van den Berg LH, Robberecht W. EPHA4 is a disease modifier of amyotrophic lateral sclerosis in animal models and in humans. *Nat Med*. 2012;18:1418-1422.

91. Takeuchi H, Kobayashi Y, Yoshihara T, Niwa J, Doyu M, Ohtsuka K, Sobue G. Hsp70 and Hsp40 improve neurite outgrowth and suppress intracytoplasmic aggregate formation in cultured neuronal cells expressing mutant SOD1. *Brain Res*. 2002;949:11–22.

92. Duan W, Guo Y, Jiang H, Yu X, Li C. MG132 enhances neurite outgrowth in neurons overexpressing mutant TAR DNAbinding protein-43 via increase of HO-1. *Brain Res*. 2011;1397:1–9.

93. Fujii R, Okabe S, Urushido T, Inoue K, Yoshimura A, Tachibana T, Nishikawa T, Hicks GG, Takumi T. The RNA binding protein TLS is translocated to dendritic spines by mGluR5 activation and regulates spine morphology. *Curr Biol.* 2005;15:587–593.

94. Janke C, Kneussel M. Tubulin post-translational modifications: encoding functions on the neuronal microtubule cytoskeleton. *Trends Neurosci.* 2010;33:362-372.

95. Romaniello R, Arrigoni F, Bassi MT, Borgatti R. Mutations in α- and β-tubulin encoding genes: Implications in brain malformations. *Brain Dev.* 2014;pii:S0387-7604(14)00137-5.

96. Bömmel H, Xie G, Rossoll W, Wiese S, Jablonka S, Boehm T, Sendtner M. Missense mutation in the tubulin-specific chaperone E (Tbce) gene in the mouse mutant progressive motor neuronopathy, a model of human motoneuron disease. *The Journal of Cell Biology* 2002;159, 563–569.

97. Jiang YM, Yamamoto M, Kobayashi Y, Yoshihara T, Liang Y, Terao S,Takeuchi H, Ishigaki S, Katsuno M, Adachi H, Niwa J, Tanaka F, Doyu M, Yoshida M, Hashizume Y, Sobue G. Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis. *Ann. Neurol*. 2005;57:236–251.

8. ACKNOWLEDGMENTS

I start by thanking the past and present colleagues at Laboratory of Neuroscience at Istituto Auxologico Italiano for their costant support and encouragement during my PhD thesis. I would like to express my special appreciation and thanks to my advisors, Professor Vincenzo Silani, Dr. Antonia Ratti and Dr. Nicola Ticozzi, who always encouraged me during my research project in these years, providing scientific guidance and financial support.

I also thank my PhD co-supervisor, Professor Giacomo Comi, for his advice and suggestions in my thesis work.

A special thank goes to Professor John Landers and his research group at University of Massachussetts, Boston, for the significative and productive collaborations that led to the results discussed in this dissertation. I would also like to acknowledge all the other research groups, in particular the clinical Centers participating to the Italian SLAGEN Consortium, that I have had the pleasure to collaborate with.

In my training period at Institute of Psychiatry-King's College London, London, I really appreciated the support of Dr. Isabella Fogh and the members of Professor John Powell's group.

Finally, special thanks go to all ALS patients and families who participated in these studies.

Research support was provided by the Italian Ministry of Health (RF 2009 – 1473856), by AriSLA - The Agency for Research on Amyotrophic Lateral Sclerosis (grants EXOMEFALS 2010 and NOVALS 2012 co-financed with the contribution of 5x1000 – Healthcare Research support of the Ministry of Health), and by IRCCS Istituto Auxologico Italiano.

9. SCIENTIFIC PRODUCTION

Fogh I, Ratti A, Gellera C, Lin K, **Tiloca C**, Moskvina V, Corrado L, Sorarù G, Cereda C, Corti S, Gentilini D, Calini D, Castellotti B, Mazzini L, Querin G, et al. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum Mol Genet*. 2014; 23:2220-31.

Ticozzi N, **Tiloca C**, Calini D, Gagliardi S, Altieri A, Colombrita C, Cereda C, Ratti A, Pezzoli G, Borroni B, Goldwurm S, Padovani A, Silani V. C9orf72 repeat expansions are restricted to the ALS-FTD spectrum. *Neurobiol Aging* 2014; 35:936.e13-7.

Tiloca C, Ticozzi N, Pensato V, Corrado L, Del Bo R, Bertolin C, Fenoglio C, Gagliardi S, Calini D, Lauria G, Castellotti B, Bagarotti A, Corti S, Galimberti D, Cagnin A, Gabelli C, Ranieri M, Ceroni M, Siciliano G, Mazzini L, Cereda C, Scarpini E, Sorarù G, Comi GP, D'Alfonso S, Gellera C, Ratti A, Landers JE, Silani V; SLAGEN Consortium. Screening of the PFN1 gene in sporadic amyotrophic lateral sclerosis and in frontotemporal dementia. *Neurobiol Aging*. 2013;34:1517.e9-10.

Gellera C, **Tiloca C**, Del Bo R, Corrado L, Pensato V, Agostini J, Cereda C, Ratti A, Castellotti B, Corti S, Bagarotti A, Cagnin A, Milani P, Gabelli C, Riboldi G, Mazzini L, Sorarù G, D'Alfonso S, Taroni F, Comi GP, Ticozzi N, Silani V; SLAGEN Consortium. Ubiquilin 2 mutations in Italian patients with amyotrophic lateral sclerosis and frontotemporal dementia. *J Neurol Neurosurg Psychiatry* 2013; 84:183-187.

Ratti A, Corrado L, Castellotti B, Del Bo R, Fogh I, Cereda C, **Tiloca C**, D'Ascenzo C, Bagarotti A, Pensato V, Ranieri M, Gagliardi S, Calini D, Mazzini L, Taroni F, Corti S, Ceroni M, Oggioni GD, Lin K, Powell JF, Sorarù G, Ticozzi N, Comi GP, D'Alfonso S, Gellera C, Silani V; SLAGEN Consortium.C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. *Neurobiol Aging*. 2012; 33:2528.e7-14.

Wu CH, Fallini C, Ticozzi N, Keagle PJ, Sapp PC, Piotrowska K, Lowe P, Koppers M, McKenna Yasek D, Baron DM, Kost JE, Gonzalez-Perez P, Fox AD, Adams J, Taroni F, **Tiloca C**, Leclerc AL, Chafe SC, Mangroo D, Moore MJ, Zitzewitz JA, Xu ZS, van den Berg LH, Glass JD, Siciliano G, Cirulli ET, Goldstein DB, Salachas F, Meininger V, Rossoll W, Ratti A, Gellera C, Bosco DA, Bassell GJ, Silani V, Drory VE, Brown RH Jr, Landers JE. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature* 2012; 488:499-503.

Tiloca C, Ratti A, Pensato V, Castucci A, Sorarù G, Del Bo R, Corrado L, Cereda C, D'Ascenzo C, Comi GP, Mazzini L, Castellotti B, Ticozzi N, Gellera C, Silani V; SLAGEN Consortium. Mutational analysis of VCP gene in familial amyotrophic lateral sclerosis. *Neurobiol Aging* 2012; 33:630.e1-2