ABSTRACT:

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disorder that results in loss of the upper and lower motor neurons from motor cortex, brainstem and spinal cord. Whilst the majority of cases are sporadic, around 10% show familial inheritance. ALS is usually inherited in an autosomal dominant manner, though autosomal recessive and X-linked inheritance do occur. To date, 24 of the genes at 26 loci have been identified; these include loci linked to ALS as well as to FTD-ALS, where family pedigrees contain individuals with frontotemporal dementia with/without ALS. The most commonly established genetic causes of FALS to date are the presence of a hexanucleotide repeat expansion in the C9ORF72 gene (39.3% FALS) and mutation of SOD1, TARDBP and FUS, with frequencies of 12-23.5%, 5% and 4.1% respectively. However, with the increasing use of next generation sequencing of small family pedigrees, this has led to an increasing number of genes associated with ALS. This review provides a comprehensive review on the genetics of ALS and an update of the pathogenic mechanisms associated with these genes. Commonly implicated pathways have been established, including RNA processing, the protein degradation pathways of autophagy and ubiquitin-proteasome-system as well as protein trafficking and cytoskeletal function. Elucidating the role genetics plays in both FALS and SALS is essential for understanding the subsequent cellular dysregulation that leads to motor neuron loss, in order to develop future effective therapeutic strategies.

KEYWORDS:

Genetics, ALS, C9ORF72, TDP-43, RNA processing, autophagy

RUNNING HEADER:

Genetics of ALS
INTRODUCTION

Amyotrophic lateral sclerosis is a progressive neurodegenerative disorder with an incidence of 2-3 per 100,000 and a lifetime risk of 1 per 400 individuals. Usually with an adult onset, initial clinical symptoms such as loss of dexterity in the fingers or a mild limp in limb onset ALS, or slurring of speech in bulbar onset ALS, are caused by loss of the upper motor neurones in the motor cortex and brainstem and lower motor neurones in the spinal cord. Overtime, the progressive nature of the disease is associated with further muscle wasting, weight loss, fasciculations and eventually death due to respiratory failure, 32 months following symptom onset. During disease progression, cognitive impairment may develop in up to 40% of patients and approximately 5% will go on to develop frontotemporal dementia (FTD).

Whilst the majority of cases are sporadic (SALS) with no familial history of disease, around 10% of cases are familial (FALS) and are clinically indistinguishable from SALS cases. Generally, in adult onset ALS, the disease is inherited as an autosomal dominant (AD) trait, though rare cases of juvenile ALS are more commonly associated with autosomal recessive (AR) inheritance. However, there are also instances of AR inheritance of AD genes in specific populations (SOD1 in Scandinavia and FUS in Cape Verde). In addition, there is evidence of reduced penetrance of disease associated mutations (including p.I114T SOD1 and G4C2 C9ORF72) as well as oligogenic inheritance, illustrating that ALS is a highly complex genetic disorder.

The first genetic cause of ALS was identified in 1993 through linkage analysis to the Chr21 marker DS21S223. Subsequent analysis of the nearby gene, SOD1, identified multiple pathogenic mutations in these FALS families. Over the following 16 years, linkage analysis and
candidate gene sequencing of ALS families have identified further genes associated with autosomal dominant or autosomal recessive ALS and one instance of X-linked inheritance (Table 1). With the advent of next generation sequencing (NGS), whole exome sequencing (WES) has allowed an exponential increase in the identification of disease associated genes (Figure 1). Following the first use of WES to identify that the VCP gene was associated with disease, WES has been used on many FALS samples and assisted in identifying seven further genes in the last three years.

To date, over 22 ALS and 4 ALS+FTD (FTDALS) loci have been established, with the causative genes identified in the majority of cases. This review will firstly summarise the current insights that have been gained from the four most common causes of FALS: 

- **SOD1**
- **TARDBP**
- **FUS**
- **C9ORF72**

These genes have highlighted the roles of oxidative stress and RNA processing as contributory pathogenic mechanisms in ALS. In addition, rarer genetic variants have implicated additional biological pathways such as the ubiquitin-proteasome system (UPS), protein trafficking and impaired cytoskeletal function.

[Please note, for the purposes of this review, we have used the numbering of loci as described in the Online Mendelian Inheritance in Man (OMIM) Phenotypic Series for ALS (PS105400) and FTDALS (PS105550).]

**ALS1: Cu-Zn Superoxide Dismutase (SOD1)**

Mutation of Cu-Zn superoxide dismutase 1 (SOD1) was the first described genetic cause for familial ALS. The majority of SOD1 mutations are AD, with a highly penetrant pattern of
inheritance, and are primarily associated with limb onset ALS. An exception to this rule is the
D90A mutation, predominantly found in Scandinavian populations, where it is inherited in an AR
manner. Frequency of SOD1 mutations varies depending on populations, from 23.5% in
Scandinavia to 12% in Germany; mutations have also been identified in apparently sporadic
cases. The ALSoD database (http://alsod.iop.kcl.ac.uk) reports that there are 183 mutations in
SOD1 associated with disease (accessed Nov 2015), the majority of which are point mutations.

Given that SOD1 encodes a 153 amino acid protein, this number is remarkable, with the
mutations distributed throughout the gene and impacting upon a variety of domains within the
protein. This is in contrast to some of the other ALS-associated mutations, which are more often
located within a particular motif of the protein product, particularly as it is unclear whether all of
the reported SOD1 “mutations” are indeed pathogenic.

The multiple mutations throughout the protein have also resulted in challenges determining how
they are responsible for the disease phenotype. SOD1 is a ubiquitously expressed antioxidant
protein, which catalyses free radical superoxide to hydrogen peroxide and oxygen. As the
majority of mutant proteins retain this enzymatic function, the pathogenicity is proposed to act
through a toxic gain of function, although the precise nature of this toxicity remains to be
determined. Many mutually compatible pathogenic mechanisms have been proposed including
oxidative stress, excitotoxicity, protein aggregation, neuroinflammation, apoptosis, mitochondrial
dysfunction, axonal transport dysregulation, endoplasmic reticulum (ER) stress. Mutant SOD1
proteins (mtSOD1) show variable states of metallation and disulphide bond formation, which
leads to the ability of the demetalled and unfolded apoform to enter the intermembrane space of
the mitochondrion thereby causing mitochondrial dysfunction. In addition, demetallation leads to
increased instability and mtSOD1 shows a higher aggregation propensity than wild type SOD1.
(wtSOD1). More recently, mtSOD1, along with misfolded wtSOD1 has been shown to move from cell to cell and initiate a prion-like seeded aggregation of SOD1. Whilst initial work demonstrated the propagation of misfolded protein in cell culture models, spinal homogenates from paralysed mutant G93A SOD1 mice injected into 6mth old G85R-SOD1:YFP mice (who don’t usually get disease till 20mths) produced a progressive motor neuron disease within 3mths.

Normally, misfolded proteins are removed from the cell via the UPS. However in SOD1-ALS, and also in SALS, the UPS has been shown to be impaired. In addition, mahogunin ring finger 1, an E3 ubiquitin ligase, which catalyses monoubiquitination of proteins and marks them for degradation via a UPS-independent mechanism, has been shown to be decreased in the G93A mouse model. Interestingly, however, overexpression of this protein reduced SOD1 toxicity by suppressing the aggregation of SOD1. Thus, therapeutic strategies for ALS include increasing clearance of misfolded SOD1 and the heat-shock protein inducer, arimoclomol, is one such drug currently under investigation.

Whilst initially oxidative stress was thought to be one of the primary mechanisms of mutant SOD1, the continued research on SOD1 pathogenic mechanisms has implicated UPS, protein aggregation and degradation, as well as other aspects of protein trafficking. These pathways are also implicated by the discovery of additional FALS genes (see “Protein Trafficking and Degradation Related Genes”).

**ALS10: TAR DNA Binding Protein (TARDBP)**
The transactive response DNA-binding protein 43 (TDP-43) is encoded by TARDBP gene on chr1p36.22\(^{16}\). TARDBP is responsible for 4-5% of FALS and nearly 1% of SALS\(^{17}\). Mutations in TARDBP are inherited in an AD manner and are associated with a classic ALS clinical phenotype. TARDBP encodes several isoforms of which TDP-43 is considered the most prevalent. TDP-43 is a heterogeneous nuclear ribonucleoprotein (hnRNP) with a nuclear localisation signal (NLS) and nuclear export signal (NES) which allows shuttling of the protein between the nucleus and cytoplasm. The TDP-43 protein contains three further domains, two RNA recognition motifs (RRM1 and RRM2), involved in RNA and DNA binding, and a glycine rich domain (GRD), which is essential for interactions with other proteins and is the location of the majority of mutations\(^{18,19}\).

TDP-43 was initially identified as a transcription repressor that binds to TAR DNA in human immunodeficiency virus-1 (HIV-1)\(^ {20}\). Subsequently, TDP-43 has been shown to play a role in RNA metabolism, including: RNA transcription, alternative splicing, pre-miRNA processing, RNA transport and mRNA stability\(^ {21}\). TDP-43 has the ability to auto-regulate its own gene expression by binding to the 3' untranslated region (3'UTR) of its mRNA, leading to instability and decay\(^ {22}\). TDP-43 also binds to UG rich sequences in multiple mRNA sequences to regulate splicing\(^ {23-25}\). In addition, a novel function has recently been described, with TDP-43 able to repress the splicing of non-conserved exons known as cryptic exons\(^ {26}\). Removal of TDP-43 allowed these cryptic exons to be incorporated into mRNA sequences, which subsequently disrupted translation and induced nonsense mediated decay (NMD). Finally, TDP-43 is also known to be a component of stress granules (SGs), although it is unclear whether this role contributes to neurodegeneration\(^ {27}\).
TDP-43 is a prominent protein in the characteristic ubiquitinated cytoplasmic inclusions found in ALS and FTD patients. Approximately 97% of FALS and SALS patients are positive for TDP-43 inclusions in the motor cortex and spinal cord, thereby establishing TDP-43 as a major protein signature for disease, not just those carrying \textit{TARDBP} mutations. The loss of TDP-43 nuclear localization in ALS is well documented and resulting splicing deficits in ALS cellular and animal models as well as in patient samples have been reported. In addition to a loss of nuclear function, a cytoplasmic gain of function may also contribute to neurodegeneration. A mouse model with a mutation in the NLS of human \textit{TARDBP}, thereby limiting TDP-43 to the cytoplasm, showed increased expression of transcription related and chromatin assembly genes as well as histone 3' UTR processing genes. Importantly, these transcriptional changes were not seen when an antisense oligomer was added to knockdown TDP-43 expression, thereby supporting a cytoplasmic toxic gain of function. Finally, similar to the prion-like propagation of disease described in SOD-ALS, there is also evidence that wTDP-43 oligomers can spread horizontally from cell to cell via microvesicles, including from ALS patient brain lysates, as well as vertically along axons. Thus, reducing the aggregation of these mutant proteins is becoming a more widely applicable therapeutic strategy.

**ALS6: Fused in Sarcoma (FUS)**

The FUS gene on chr16p11.2 was first identified as a fusion oncogene in liposarcoma. FUS belongs to the FET protein family and has been shown to be an hnRNP due to its involvement in transcription process, transport, trafficking, alternative splicing and miRNA processing. Similar to TDP-43, it is also present in SGs. Structurally, FUS is comprised of 526 amino acids which form an N-terminal domain rich in glutamine-glycine-serine-tyrosine (QGSY), three arginine-glycine-
glycine (RGG) rich domains, an RRM and a zinc finger motif as well as NES and NLS which enables nucleocytoplasmic shuttling of the protein.

Mutations in the FUS gene were initially identified in an AR Cape Verde family, although subsequent screening established FUS to be causative also in AD ALS. FUS mutations account for about 4% FALS and 1% of SALS, with the majority of mutations located either within exons 3-6, encoding the QGSY-rich and first RGG region, or in exons 12-15, which encode zinc finger domain, the other 2 RGG domains and NLS. Whilst these in the C-terminal have been shown to be functional, those in exons 3-6 are more commonly found in SALS or do not always segregate with disease, suggesting incomplete penetrance or non-pathogenic variations.

Previously, depletion of RNA polymerase II (RNAP II) from the nucleus had been shown to lead to an increase of cytoplasmic FUS, suggesting FUS had a role in transcription. It has subsequently been shown that FUS mediates the interaction between RNAP II and the splicing factor U1 snRNP, thereby coupling transcription to splicing. Mutations in FUS lead to mislocalisation of both FUS and U1 snRNP to the cytoplasm and other RNA binding proteins, including SMN1, hnRNP A1 and A2 also co-localise in mtFUS aggregates. The consequence of these mtFUS interactions includes dysregulated splicing and an increased binding of FUS with SMN, leading to a reduction in Gem bodies, thereby representing both a loss and gain of function by the mtFUS.

FUS mutations may also confer pathogenicity via additional interactions. FUS has been shown to bind mRNAs and facilitate their transport down dendrites and subsequently has been shown to bind to the polyA tail of AMPA receptor GluA1, regulating its stability, with the loss of FUS leading
to a reduction of GluA4. FUS has also been shown to translocate to the mitochondria through interaction with the mitochondrial chaperone heat shock protein 60 (HSP60) leading to mitochondrial damage. Finally, mtFUS interacts with Pur-alpha in SGs and increases phosphorylation of the elongation initiation factor 2-alpha (eIF2-alpha), consequently inhibiting protein synthesis. However, the contribution of each of these interactions on disease pathogenesis remains to be determined.

**FTD-ALS1: C9ORF72 (C9ORF72)**

The most common cause of FALS to date is the expansion of an intronic GGGGCC repeat in C9ORF72. The region was originally identified through genome wide association studies of SALS cases, as well as within the Finnish ALS population; whilst initially sequencing of the gene failed to identify any point mutations, targeted NGS of the region established an intronic repeat region between the non-coding exons 1a and 1b. Whilst healthy controls most commonly have less than 10 hexanucleotide repeats, individuals with ALS usually carry 400-2000 repeats. The repeat expansion has been identified in 37.6% of FALS and 6.3% of SALS, as well as in up to 25.1% of familial FTD cases. Therefore, it is not surprising that the most significant clinical phenotype associated with this genetic subtype is an increased incidence of a family history of FTD. In addition, there is evidence that there are more bulbar-onset cases associated with C9ORF72-related ALS (up to 44%, compared to 25-26% in non-C9ORF72 ALS), with some studies also reporting an earlier age of onset (by 1.8-5.0yrs) and shorter disease duration (by 5.7-12.0mths).
The function of the C9ORF72 gene is currently being determined, though structural analysis has established there is similarity to "differentially expressed in normal and neoplastic (DENN)-like proteins, which are GDP/GTP exchange factors regulating Rab-GTPases involved in vesicular trafficking." Further work demonstrated how C9ORF72 colocalised with Rab proteins involved in autophagy and endosomal trafficking. Whilst the function is being established, several hypotheses have been proposed to explain how the intronic hexanucleotide repeat may cause neurodegeneration: 1) Haploinsufficiency, 2) RNA toxicity and 3) Dipeptide repeat (DPR) protein toxicity.

**Haploinsufficiency:** Lower levels of C9ORF72 transcript were seen in patients with the repeat expansion compared to controls and the haploinsufficiency hypothesis was supported when knockdown of the C9orf72 homolog in zebrafish caused an axonal degeneration. In contrast, in a conditional C9orf72 knockout mouse, where C9orf72 was specifically ablated in neuronal cells, there was no evidence of a neurodegenerative phenotype. However, a systematic investigation of the expression levels of the three C9ORF72 transcripts (variant 1 = exon 1a, 2-5; variant 2 = exon 1b, 2-11; variant 3 = exon 1a, 2-11) demonstrated significantly reduced expression of variants 1 and 2 in cerebellum and frontal cortex of C9ORF72 expansion carriers, with a correlation between a higher level of expression of variant 1 and survival. This suggests that antisense oligomer strategies should avoid reducing C9ORF72 expression levels.

**RNA Toxicity:** RNA foci were identified located primarily in the nucleus and occasionally in the cytoplasm of motor neurons and were found to be composed of both sense and antisense RNA. The presence of antisense RNA foci have been shown to be correlated with TDP-43 mislocalisation but not DPRs. The repeat sequence is thought to form G-quadruplex...
structures within the cell. Many RNA binding proteins co-localise with the RNA foci, potentially sequestering them from the cell and disrupting their RNA processing functions. This may underlie the significant dysregulation of RNA splicing that is seen in the presence of the expansion, where greater disruption is present in patients with a shorter survival. RNA foci have, however, been observed in fibroblasts from asymptomatic patients and in BAC transgenic mice containing an expanded allele, whilst RNA foci and DPRs recapitulate the neuropathology of ALS, there is no evidence of neurodegeneration. This is in contrast to a mouse expressing a 66-repeat G4C2 expansion specifically in the CNS, which exhibited neuropathological, behavioural and motor deficits at 6mths.

**DPR Proteins:** Finally, it was demonstrated that the GGGGCC repeat expansion was subject to repeat-associated non-ATG (RAN) translation. Both the sense and antisense RNA are translated, forming DPR proteins comprised of poly-GA, -GP, -GR, -PA and -PR (with poly-GP generated from both antisense and sense RNA). These DPR proteins are found aggregated within the neuronal cytoplasmic inclusions and neuronal intranuclear inclusions in the motor cortex, cerebellum, hippocampus and spinal cord, which also stain positive for ubiquitin and p62. Recently, antibodies raised against each of the DPR proteins have demonstrated that there is little correlation between DPR distribution/burden and clinical phenotype, which the authors suggest is evidence against DPR aggregation being a major pathogenic mechanism. This is in contrast to work using a Drosophila model, in which DPR expression caused neurodegeneration in the fly eye.

In summary, there appears to be growing evidence supporting some form of RNA dysregulation as a contributing mechanism to C9ORF72-ALS. However, whilst different cellular and animal
models using different constructs are generating conflicting results on the contribution of each of
the three hypotheses, the precise mechanism(s) still remain to be fully elucidated. Disruption of
C9ORF72 protein function in endosomal trafficking may also be a contributory factor.

OTHER RNA PROCESSING GENES ASSOCIATED WITH ALS

Prior to the identification of TARDBP and FUS as ALS-associated genes, several RNA
processing genes had been already been implicated in ALS: angiogenin (ANG) and senataxin
(SETX). Subsequently, mutations in hnRNPA1 and matrin 3 (MATR3) have been identified
through WES and ataxin 2 (ATXN2) was identified as a risk factor.

ALS9: Angiogenin (ANG): Following the identification of the ANG single nucleotide
polymorphism (SNP) rs11701 as over-represented in ALS cases from Scotland and Ireland,
screening of ANG identified 7 missense mutations in 15 ALS cases, of which 4 were FALS
(1.54%) and 11 were SALS (0.80%) ANG is a member of the pancreatic ribonuclease
superfamily and is neuroprotective, whilst in mtANG this ability is impaired Whilst multiple
mutations have been identified, p.K17I has not always shown disease segregation. However, a
meta-analysis has demonstrated that Caucasian individuals carrying this allele have a 1.65
greater risk of ALS, which is increased to 10-fold in FALS ANG has also been shown to induce
the assembly of SGs Interestingly, this induction is inhibited by G-quadruplex structures, which
are formed by the G4C2 C9ORF72 expanded repeat, thereby establishing a link between
C9ORF72 and ANG.
ALS4: Senataxin (SETX): Mutations in SETX are associated with the juvenile onset of ALS with distal muscle weakness and absence of bulbar or sensory symptoms. Patients typically have a long and slow disease progression with a relatively normal life span. Rare, AD mutations in SETX occur in ALS, while recessive SETX mutations are associated with ataxia-oculomotor apraxia-2 (AOA2). The mechanisms by which SETX variants lead to ALS is unknown; however, SETX encodes a DNA/RNA helicase protein proposed to play a role in DNA repair in response to oxidative stress. SETX also interacts with RNA processing proteins regulating transcription and pre-mRNA processing suggesting the cause of motor neuron degeneration through SETX mutations is as a result of abnormal RNA processing.

ALS13: Ataxin 2 (ATXN2): More than 36 repeats of CAG within ATXN2 causes spinocerebellar ataxia 2; however, intermediate repeats of 27-33 were found to strongly associate with ALS having established that ATXN2 modifies TDP-43 toxicity in yeast. ATXN2 is an RNA binding protein that is involved in RNA processing and localised to the ER, Golgi apparatus and SGs; ATXN2 also interacts with FUS and intermediate expansions exacerbate the FUS mutant phenotype in cellular models. A recent meta-analysis of over 6000 ALS and 7000 controls has identified that repeat lengths of 25-28 were actually protective, whereas the significant risk was associated with CAG repeats of 31-33. This finding is supported by an Italian study, where additionally, <31 repeats were associated with spinal onset ALS and a shorter survival.

ALS20: Heterogeneous nuclear ribonucleoprotein A1 (hnRNPA1): Following the identification of hnRNPA1 and hnRNPA2B1 mutations as causative in multisystem proteinopathy families using exome sequencing, these genes were specifically analysed in 212 FALS for which exome sequencing was available. A single case was identified with a mutation in hnRNPA1.
Interestingly, hnRNPA1 and A2/B1 are known interacting partners of TDP-43 and hnRNPA1 also interacts with ubiquilin-2. In ALS motor neurons, there is a loss of intense hnRNPA1 nuclear staining, which also correlates with nuclear loss of TDP-43, although hnRNPA1 was not seen to co-localise with TDP-43 in the skein-like inclusions. However, screening of 113 Italian FALS as well as 135 Dutch FALS and 1084 Dutch SALS failed to find any hnRNPA1 mutations, suggesting this is a very rare cause of FALS.

**ALS21: Matrin 3 (MATR3):** Exome sequencing of a large pedigree identified a mutation in the MATR3 gene; a previous family carrying a mutation in MATR3, originally diagnosed with AD distal, asymmetrical myopathy with vocal cord paralysis were re-assessed and re-diagnosed as having ALS. Further screening of Italian and British ALS cases identified a further 2 mutations; one FALS and one SALS. MATR3 is an RNA/DNA binding protein that interacts with TDP-43; whilst the p.S85C mutation enhances this interaction, two other mutations, p.F115C and p.T22A, do not. However, this difference may underlie the slow progression of the disease in the family carrying the p.S85C mutation. Whilst no further mutations were found in 372 FALS cases from France, Taiwan, Australia and French Canada, 4 mutations were found in apparent SALS cases (3 in French-Canadians and in 1 Taiwanese).

**FET Family Genes:** The TATA box-binding protein-associated factor 15 (TAF15) and Ewing sarcoma breakpoint region 1 (EWSR1) are RNA binding proteins and form the FET family of proteins along with FUS. All three contain prion-like domains, a feature used to rank potential RNA binding proteins as being involved in ALS following a functional yeast screen. Screening of TAF15 identified 5 missense variants in 1262 ALS cases, whilst screening EWSR1 identified 2 potential mutations among 817 ALS cases. Whilst these variants were absent from controls,
they were identified in SALS patients, so segregation could not be demonstrated. However, both TAF15 and EWSR1 proteins show cytoplasmic mislocalisation in SALS. Recently, whole genome sequencing identified an EWSR1 mutation in a set of monozygote twins discordant for ALS, suggesting additional factors may influence disease.

PROTEIN TRAFFICKING AND DEGRADATION RELATED GENES

From the identification of the first gene for AR ALS, ALS2, combined with the presence of characteristic ubiquitinated inclusions in ALS motor neurons, dysregulation of protein trafficking and protein degradation have been implicated in the disease process. Mutations in genes involved in endosomal transport include alsin (ALS2), vesicle associated membrane protein [VAMP] associated protein B (VAPB), chromatin modifying protein 2B (CHMP2B) and phosphoinositide 5-phosphatase (FIG4), those involved in the UPS include ubiquilin 2 (UBQLN2), sequestosome 1 (SQSTM1) and sigma non-opioid intracellular receptor 1 (SIGMAR1) and autophagy is primarily implicated by mutations in optineurin (OPTN), valosin containing protein (VCP) and tank-binding kinase 1 (TBK1). There is some overlap between these three biological pathways, which have also been implicated in SOD1 and C9ORF72-related ALS.

ALS2: Alsin (ALS2): The ALS2 gene was originally identified through linkage analysis in consanguineous families from Tunisia and Saudi Arabia. The majority of the mutations lead to protein truncations, leading to a proposed loss of function. Alsin is thought to play a role activating Rab5 GTPases. Rab5 is essential for endosomal trafficking and in alsin knockout mice, neurons showed increased endosomal fusion and degradation but reduced mobility. One of the
endosome components is the AMPA receptor GluR2, levels of which are reduced in alsin knockout mice.

**ALS8: Vesicle associated membrane protein [VAMP] associated protein B (VAPB):** Linkage analysis of a large Brazilian family first established VAPB as a causative gene for ALS and the p.P56S mutation has been identified in multiple Brazilian pedigrees, indicative of a common founder. Additional mutations have been reported, though not all of the variations found segregated with disease. VAPB is a type II integral ER membrane protein, involved in intracellular trafficking and the unfolded protein response (UPR) as well as regulating ER-mitochondria interactions. However, the p.P56S mutant protein is unable to initiate the UPR, altered calcium uptake into the mitochondria and disrupted anterograde axonal transport of mitochondria.

**ALS17: Chromatin modifying protein 2B (CHMP2B):** Mutations in CHMP2B were initially identified in two probable FALS and a further 3 SALS cases; the majority had a predominant lower motor neuron phenotype. CHMP2B is a component of the ESCRT-III endosomal trafficking system, sorting cargos into multivesicular bodies (MVBs). More recently 4 novel mutations were identified in apparently sporadic ALS cases, and these were located in the domain required to form the MVB. In cellular models, mutant CHMP2B led to the formation of large vacuoles, and an increase in autophagy marker LC3-II, implicating dysregulation of autophagy as a mechanism in ALS.

**ALS11: Phosphoinositide 5-phosphatase (FIG4):** Mutations in FIG4 were originally identified as causative in Charcot-Marie-Tooth disease type 4J, though one family had a clinical phenotype
resembling ALS. Screening of FALS and SALS cases identified 9 variants, with 6 showing impaired function in yeast. FIG4, also known as SAC3, regulates PI(3,5)P$_2$ levels, and thereby controls retrograde trafficking of endosomal vesicles to the Golgi. The mutant proteins showed loss of phosphatase activity, mislocalisation and inability to bind to the PI(3,5)P$_2$ complex. Further screening in Italian and Taiwanese populations failed to find any novel variants, though only 80 SALS and 15 FALS were screened in each study. No pathological assessment was available on the mutation carriers, however, FIG4 was not shown to be mislocalized in SALS.

**ALS15: Ubiquilin 2 (UBQLN2):** UBQLN2 was identified by linkage analysis in a large multigenerational family. Screening of additional FALS cases with no male to male transmission found 4 further mutations; these were all located within the PXX repeat region of the protein. Additional screening identified further variants adjacent or within the PXX repeat region. Mutations have been shown to disrupt the protein degradation pathway through defective binding to the proteasome and causing mislocalisation of OPTN from Rab-11 positive endosomal vesicles, as well as potentially impairing RNA metabolism, through loss of binding of UBQLN2 to hnRNP proteins, including hnRNP.

**FTDALS3: Sequestosome 1 (SQSTM1):** SQSTM1 or p62 is a ubiquitin binding protein that plays a role in protein degradation via the proteasome and autophagy, and is found within the characteristic ubiquitinated inclusions in ALS patients. Screening of this gene found multiple mutations in both FALS and SALS cases. Further mutations were found in ALS patients, some in association with Paget disease of bone, which is also known to be caused by SQSTM1 mutations. In a zebrafish model, where endogenous SQSTM1 was knocked-down, the fish showed behavioural and axonal abnormalities, as well as disrupted autophagy, as demonstrated.
by increased mTOR levels. Human SQSTM1 was able to rescue the phenotype, but the frequently found mutation, p.P392L, was unable to do so.

**ALS16: Sigma non-opoid intracellular receptor 1 (SIGMAR1):** Initially 3'UTR variants were identified within several AD FTD-ALS or FTD families and suggested pathogenicity occurred through alteration of mRNA stability. However, using homozygosity mapping, a missense mutation in SIGMAR1 was subsequently found to segregate in a large consanguineous family with AR juvenile ALS. SIGMAR1 is an ER chaperone, a subunit of the ligand-regulated potassium channel and enables mitochondrial calcium transport via the IP3 receptor; mutation of SIGMAR1 causes the formation of cytoplasmic aggregations, a reduction in ATP production and subsequent decrease in proteasome activity. However, whether SIGMAR1 contributes to AD ALS is yet to be fully elucidated.

**ALS12: Optineurin (OPTN):** Mutations in OPTN were originally identified through homozygosity mapping of consanguineous Japanese AR ALS families, which identified a homozygous exonic deletion and a homozygous nonsense mutation. Further screening of FALS cases identified two AD families heterozygous for a missense mutation. OPTN mediates its function through protein-protein interactions; it binds to ubiquitin and UBQLN2, is an autophagy receptor (facilitating the recruitment of cargos to autophagosomes), is required for Golgi organisation (as demonstrated by the Golgi fragmentation seen in post-mortem spinal motor neurons) and also regulates NFkB signalling. Subsequent screening has identified additional heterozygote mutations, including in SALS cases as well as in AR ALS cases.
ALS14: Valosin containing protein (VCP): Exome sequencing of a four-generation Italian family initially suggested mutation of VCP as causative in ALS. Subsequently, a further 4 variants were identified in FALS cases, thereby providing further evidence that VCP is associated with ALS. VCP is an AAA+-ATPase involved in a variety of cellular functions including mediating the proteasomal degradation of ubiquitinated protein in multimeric complexes and the targeting of substrates to autophagosomes. Whilst screening failed to find any VCP mutations in some populations potential mutations were found in others and also in SALS cases. VCP mutations are also associated with IBMFD and patient fibroblasts have shown mitochondrial uncoupling and a reduction in ATP production, a feature also seen with SIGMAR1 mutations.

FTDALS4: TANK-binding kinase (TBK1): Mutations in TBK1 were initially identified through exome sequencing 2874 ALS cases; dominant variants were found in 1.097% of cases and loss of function mutations in 0.382%. This was shortly followed by a second paper in which sequencing of 252 FALS cases identified 9 loss of function and 4 missense mutations. Mutations have subsequently been found in both ALS, FTD-ALS and FTD cases. TBK1 has a role in both innate immunity, NFkB signalling, as well as in autophagy. TBK1 binds and phosphorylates ALS-related proteins OPTN and SQSTM1, whereas TBK1 mutants have been shown to no longer bind OPTN.

IMPAIRED AXONAL TRANSPORT AND CYTOSKELETAL DYSFUNCTION

Neurons are extremely large cells which require transport of organelles, proteins and RNA from the cell body down the axons. Molecular motors such as kinesins and dynein guide these cargos to microtubules to mediate anterograde and retrograde transport, respectively. Whilst mutation of
the p150 dynactin subunit in mice generated a neurodegenerative phenotype, mutations in this gene were not found in human ALS. However, exome sequencing has identified several cytoskeletal genes with mutations reported to be causative in ALS.

**ALS5: Spatacsin (SPG11):** WES of two affected siblings from a non-consanguineous family diagnosed with autosomal recessive juvenile onset ALS (ARJALS) identified only one gene, SPG11, in which variants were found in a compound heterozygous state. The involvement of the hereditary spastic paraparesis gene, normally associated with HSP with thin corpus callosum, had previously been implicated by a candidate gene screening of SPG11, in which mutations in 10 families with ARJALS were identified. Whilst the exact function of the protein is unknown, iPSC-derived neuronal cells with SPG11 mutations demonstrated the protein co-localised with the cytoskeleton, and mutations caused axonal instability and impaired axonal transport.

**ALS18: Profilin 1 (PFN1):** Two multi-generational ALS families were determined to carry mutations in the PFN1 gene following WES. Extending the screen to additional FALS cases, identified a further 3 mutations in 5 FALS cases as well as a p.E117G variant, that was identified at very low frequency in controls. Further screening of ALS cases identified additional mutations as well as the variant. A meta-analysis subsequently determined that the p.E117G was associated with ALS and proposed this variant as a risk factor. The function of PFN1 is to convert monomeric actin to filamentous actin and it is also found localised to SGs. It has been demonstrated that PFN1 mutations destabilise the protein, thereby leading to a loss of function, whilst the mutant protein is misfolded, thereby leading to a gain of function through aberrant protein interactions. However, the effect of the mutant protein on actin formation and SG dynamics is yet to be elucidated.
ALS22: Tubulin alpha 4A (TUBA4A): WES of 363 FALS index cases followed by rare variant burden identified 5 ALS cases with rare variants in TUBA4A; these included 4 missense mutations and 1 nonsense mutation, all of which were encoded in exon 4, in highly conserved amino acids, and absent from 4300 EVS controls. Whilst further sequencing of ALS cases only identified one further variant, functional studies demonstrated that the p.W407X nonsense mutant failed to localise to the microtubules, instead forming cytoplasmic inclusions, leading to disrupted microtubule assembly and stability, through a dominant-negative mechanism. Subsequent screening in the Chinese ALS population, failed to identify any variants; data from other populations will undoubtedly emerge as further WES experiments are completed.

Intermediate Filament Variants: Cytoskeletal dysfunction is further implicated in the pathogenesis of ALS through rare variants being identified in intermediate filament genes. Neurofilaments (light, medium and heavy) are major structural components of the neuronal cytoskeleton and are present within the characteristic ubiquitinated protein inclusions. Candidate gene screening identified rare insertion/deletion variants in the KSP repeat domains of neurofilament heavy (NEFH) gene in SALS cases, whilst a single frameshift mutation has been identified in peripherin (PRPH1). However, the absence of mutations in known familial cases and ability to show segregation with disease has reduced the creditability of these genes as ALS loci.

ADDITIONAL LOCI
Several additional ALS loci have been identified. Two loci are yet to have their associated genes identified: ALS3 on chr18q21 and ALS7 on chr20q13. A further two genes have been identified in FALS pedigrees, though their functional effects are currently predicted to disrupt neuronal development and mitochondrial function.

**ALS19: Erb-b2 receptor tyrosine kinase 4 (ERBB4):** Whole genome sequencing (WGS) of a Japanese AD ALS family identified a missense mutation in ERBB4. Additional screening identified the same mutation in an unrelated Canadian family and a further mutation in a SALS case. ERBB4 is a receptor tyrosine kinase which is activated by neuregulin, resulting in autophosphorylation of the C-terminal. Mutations in ERBB4 reduced the level of autophosphorylation. ERBB4 was found to localise to C-boutons, arising from interneurons, which synapse with spinal motor neurons. Interestingly, C-boutons are not found in oculomotor neurons, which are spared in ALS, whilst increases in neuregulin levels in C-boutons increase during disease progression of the SOD1 G93A transgenic mice.

**FTDALS2: Coiled-coil helix coiled-coil helix domain containing protein 10 (CHCHD10):**

CHCHD10 was initially associated with ALS through WES of a family exhibiting clinical features including ALS, FTD, cerebellar ataxia and myopathy. This led to ALS and ALS-FTD families being screened. Several additional mutations were found, though it became evident that the p.P34S mutation was non-pathogenic, as it was also found at similar frequencies in controls. The function of CHCHD10 is unknown, it localises to the mitochondria. Fibroblasts from family members of the original pedigree showed multiple mitochondrial DNA deletions, respiratory chain defects and structurally abnormal mitochondria, suggesting CHCHD10 may have a role in the respiratory chain and/or in mitochondrial genome stability. This has been supported by
additional work by Genin and colleagues, which not only found loss of cristae and mitochondrial genome repair in patient fibroblasts, but a failure of apoptosis, due to an inability to release cytochrome C.172

**CONCLUSION**

ALS genetics is having a significant impact on our understanding of the disease and the mechanisms implicated in neurodegeneration. The majority of genes encode proteins involved in RNA processing and the protein degradation pathways, UPS and autophagy. However, neither of these or the other pathways implicated work in isolation, but impact on other cellular processes. The proposed mechanisms are mutually compatible and it is most likely that multiple dysregulated pathways contribute to the loss of motor neurons. This is clearly demonstrated by TDP-43, an RNA binding protein which is mislocalised from the nucleus, thereby causing loss of nuclear function, and which is the aggregated in the cytoplasm as a component of the characteristic ubiquitinated inclusions.

Along with multiple genetic causes, it is clear that these genes are also implicated in additional disorders, not only other neurodegenerative disorders such as FTD, HSP and ataxia, but also myopathies, Paget disease of bone and glaucoma (Table 2). The use of WES or WGS, in projects such as the 100,000 Genomes Project in the UK (www.genomicsengland.co.uk), or Project Mine among the International ALS community (www.projectmine.com), will potentially enable a greater understanding of why mutations in a gene in one family present with a specific clinical phenotype, whilst another family shows a different disease.
WES and WGS will also further our understanding of the impact of oligogenic inheritance in ALS. Whilst family pedigrees clearly show inheritance in an AD manner for classical ALS, the move away from analysing a single gene at a time has highlighted evidence of mutations in multiple ALS genes in some patients. Screening of a large ALS cohort demonstrated that 14% of FALS and 2.6% of SALS cases had more than one potential pathogenic mutation in a known ALS gene, and these cases had a significantly earlier onset of disease. This also highlights the fact that apparent SALS cases also carry genetic mutations, as has been evidenced by the identification of de novo mutations in SALS cases following WES of case and unaffected parents trios. Whilst in some cases, these may actually be rare AR mutations, additional WES and WGS sequencing of cases will allow the genetic contribution in SALS, estimated to be 61%, to be elucidated.

REFERENCES


29. Qin, H, Lim, LZ, Wei, Y & Song, J. TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc Natl Acad Sci U S A* 2014;111:18619-18624.


38. Yu, Y & Reed, R. FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proc Natl Acad Sci U S A* 2015;112:8608-8613.


93. Meltz Steinberg, K, Nicholas, TJ, Koboldt, DC, Yu, B, Mardis, E & Pamphlett, R. Whole genome analyses reveal no pathogenetic single nucleotide or structural differences...


FIGURE LEGEND

Figure 1: Gene Frequencies in ALS. Each gene is plotted against the year it was found; size of circles signifies the frequency of mutations in FALS (ALS) as listed on Table 1. Where gene frequencies were not available, for illustrative purposes, these have been given circle size equivalent to 1%.

ACKNOWLEDGEMENTS

JK and RW are funded by STRENGTH. This is an EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organisations under the aegis of JPND - Belgium, The National Funds for Scientific Research (F.R.S. FNRS); France, Agence Nationale de la Recherche (ANR); Germany, Bundesministerium für Bildung und Forschung (BMBF); Italy, Ministero della Salute; Italy, Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR); The Netherlands, The Netherlands Organisation for Health Research and Development (ZonMw); Sweden, Swedish Research Council (VR); Switzerland, Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF); United Kingdom, Medical Research Council (MRC).

JK is also are funded by SOPHIA. This is is an EU Joint Programme - Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organisations under the aegis of JPND - France, Agence Nationale de la Recherche (ANR); Germany, Bundesministerium für Bildung und Forschung (BMBF); Ireland, Health Research Board (HRB); Italy, Ministero della Salute; The Netherlands, The Netherlands Organisation for Health Research and Development (ZonMw); Poland, Narodowe Centrum
Badań i Rozwoju; Portugal, Fundação a Ciência e a Tecnologia; Spain, Ministerio de Ciencia e Innovación; Switzerland, Schweizerischer Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF); Turkey, Tübitak; United Kingdom, Medical Research Council (MRC).

JK receives funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under the EuroMOTOR project, grant agreement no 259867 and from the European Horizon 2020 research and innovation programme under grant agreement no 633413.

AAS is funded by a PhD Scholarship from the University of Dhammam, Saudi Arabia.