The current insights into the genetics of amyotrophic lateral sclerosis

Merith Bosman (S3185303)
28th of June, 2019
Neurosciences Research (WLB07067)
Supervised by Ulrich Eisel
University of Groningen
Synopsis
Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease characterized by the destruction of both upper and lower motor neurons, resulting in the inability to initiate and control movements and ultimately, in death. It is believed that multiple cellular processes are at fault, initiated by certain genetic and environmental factors. About 10% of ALS patients, referred to as familial ALS patients, has a clear genetic link to ALS since they have relatives also suffering from this disease. The vast majority known as sporadic ALS patients, however, does not. It is thought that they might possess susceptibility genes, which lead to ALS pathogenesis in combination with specific environmental risk factors. Currently, over 30 genes have been identified to be involved in familial ALS, most of which have been observed to produce proteins that normally function within RNA metabolism, vesicle trafficking, autophagy or cytoskeletal maintenance. Mutations within these genes are thought to cause loss of function or gain of a new, toxic function, leading to motor neuron degeneration. The most frequently mutated genes in fALS patients are the SOD1 gene, the TARDBP gene, the FUS gene and the C9ORF72 gene. These genes have also been found to be mutated in a select portion of sporadic ALS patients, which leads to believe that the traditional categorization between familial and sporadic disease is inconvenient. Therefore, genetic screening should be standard protocol for all ALS patients, even if no clear genetic link seems to be present.
Index

1 Introduction 4

2 Superoxide dismutase 1 (SOD1) 5
   2.1 Glutamate excitotoxicity 6
   2.2 Mitochondrial dysfunction 6
   2.3 Influence of microglial and astroglial cells 7

3 Transactive response DNA-binding protein (TARDBP) 7

4 Fused in sarcoma (FUS) 8

5 Chromosome 9 open reading frame 72 (C9ORF72) 9
   5.1 Loss of function 10
   5.2 RNA toxicity 10
   5.3 Proteotoxicity from dipeptide repeat (DPR) aggregates 11

6 Other genes linked to ALS 12

7 Discussion and conclusion 14

8 References 16
1 Introduction
Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig’s disease after the famous American baseball player who suffered from this illness, is a progressive neurodegenerative disease, characterized by the destruction of motor neurons. With a median prevalence of 4.48 patients per 100,000 individuals globally (Chio et al., 2013), ALS is the most common motor neuron disease, distinguishing itself from the others by affecting both upper as well as lower motor neurons (Hobson et al., 2016). For instance, primary lateral sclerosis involves only the upper motor neurons, while progressive muscular atrophy involves only the lower motor neurons (Hobson et al., 2016).

The upper motor neurons, located in the central nervous system, originate in the premotor cortex and the primary motor cortex of the cerebellum, also referred to as the “motor strip” of the brain (Stifani, 2014). From there, these motor neurons descend through the brainstem and spinal cord, where they form synapses with the lower motor neurons (Harvey, 2009). In the brainstem, synapse formation occurs in the cranial nerve nuclei, while in the spinal cord, the upper motor neurons synapse with the anterior horn cells, the cell bodies of the lower motor neurons. This usually happens via interneurons (Harvey, 2009). Subsequently, axons from the lower motor neurons project to a variety of muscles. The branchial motor neurons innervate the muscles of the face and neck and the somatic motor neurons are responsible for the control of the skeletal muscles to induce voluntary movements. The visceral motor neurons indirectly carry signals to the smooth muscles and glands, belonging to the autonomic nervous system. These neurons connect to ganglionic neurons of the peripheral nervous system, which do reach the target organs (Stifani, 2014).

As a result of the degeneration of the upper motor neurons, as is seen in ALS, the fine motor control of the lower motor neuron system is lost, causing spastic paresis (Stifani, 2014). Destruction of the lower motor neurons leads to denervation of the muscles, resulting in muscle atrophy paired with fasciculation, cramps and muscle weakness. Autonomic dysfunction and deficits are seen as well in a fair amount of ALS patients, which can be contributed to changes in innervation of the autonomic nervous system. Most commonly observed are urinary urgency and constipation (Piccione et al., 2015). Interestingly, motor neurons controlling the movement of the eyes persist throughout the disease. For some unknown reason, these motor neurons are resistant to degeneration (Nijssen et al., 2017).

Depending on which motor neurons are initially affected, different onset symptoms can be seen. Based on these symptoms, three different forms of ALS are identified: limb-onset, bulbar-onset and respiratory-onset ALS (Swinnen and Robberecht, 2014). Limb-onset patients first present weakness in their upper and lower limbs, among others resulting in the difficulty to walk or to use one’s hands, while patients with bulbar-onset first show dysphagia, or difficulty with swallowing, and dysarthria, slurring of speech (Wijesekera and Leigh, 2009). Respiratory-onset ALS is characterized by first influencing the respiratory muscles, complicating the patient’s ability to breathe and leading to a worse prognosis compared to the other two forms (Swinnen and Robberecht, 2014). More than 80% of ALS patients has limb-onset symptoms and about 19% of patients display bulbar-onset ALS (Li et al., 1990). Only a small minority of ALS patients suffers from the respiratory-onset disease (Swinnen and Robberecht, 2014). First symptoms usually present themselves between the age of 50 to 60 years, although some patients show signs of ALS much earlier (Swinnen and Robberecht, 2014). After the initial symptoms, the amount of affected motor neurons and muscles progressively increases, eventually leading to the inability to initiate and control all voluntary movements. Ultimately, patients die of respiratory complications, due to the decline in strength of their respiratory muscles (Lechtzin et al., 2018). On average, patients die
within 2 to 4 years after diagnosis, but overall survival time varies between a few months to decades (Chio et al., 2009).

The mechanism underlying the degeneration of the upper and lower motor neurons is not yet fully understood, although it is known that several aspects within the neurons and brain affected by ALS are altered. For example, mitochondria within the neurons do not function properly anymore and it has been observed that the synaptic glutamate concentrations are increased, resulting in glutamate excitotoxicity (Bonafede and Mariotti, 2017). Heightened levels of free radicals and oxidative damage were also found in the cerebrospinal fluid (CSF) of ALS patients (Zarei et al., 2015), implying a rise in oxidative stress in the brain. Furthermore, it is believed that neuroinflammation contributes to the motor neuron degeneration (Komine and Yamanaka, 2015). The consensus is that these alterations are initiated by both genetic as well as environmental factors. About 10% of ALS patients have first or second degree relatives also suffering from this motor neuron disease, strongly suggesting genetic involvement. Patients in this group are referred to as familial ALS (fALS) patients. The remainder of patients, known as sporadic ALS (sALS) patients, does not show an apparent genetic link, but it is thought that they might possess “susceptibility” genes, that may cause ALS in combination with certain environmental risk factors (Bonafede and Mariotti, 2017).

By identifying the genes linked to ALS, the hope is to get a better understanding of the mechanisms underlying ALS, thus facilitating the development of new therapies. At present, only a few therapies have been developed to treat ALS, but unfortunately these have been unsuccessful to cure the disease. For example, riluzole administration only prolongs the patient’s life by 3 months (Miller et al., 2012). In this review, I intend to answer which genes are currently linked to familial ALS and to discuss how they affect cell function, causing degeneration of upper and lower motor neurons. By engaging in these disrupted pathways, perhaps new therapeutic strategies can be developed.

2 Superoxidase dismutase 1 (SOD1)

In 1993, the first gene was associated with ALS pathogenesis: the SOD1 gene, located on chromosome 21q12.1. Using linkage analysis in several pedigrees with fALS, researchers found multiple genetic mutations on this locus, leading to its connection with the motor neuron disease (Rosen et al., 1993). About 14.8% of fALS patients from European descent and 30.0% of fALS patients from Asian descent have a mutated SOD1 gene, which makes it one of the major causative genes for ALS (Zou et al., 2017).

The SOD1 gene consists of five small exons separated by four introns and encodes for the antioxidant enzyme superoxidase dismutase 1 (SOD1), a homodimer of two polypeptides, each 153 amino acids long and with a mass of 16 kDa (Andersen, 2006; Zelko et al., 2002). Both polypeptides include binding sites for one zinc ion and for one copper ion, necessary to stabilize the homodimeric structure of the enzyme or to catalyze SOD1 activity, respectively (Alemasov et al., 2018). The eight antiparallel β-sheets of each polypeptide contribute to the dimer stability as well, due to the strong hydrophobic interactions between the β-strands (Andersen, 2006). The main responsibility of this protein is to eliminate superoxide free radicals (O2•−), which could inflict oxidative damage on cells. In the so-called dismutation reaction, superoxide free radicals are reduced to oxygen (O2) and the less reactive hydrogen peroxide (H2O2), which, in turn, is reduced to water (H2O) by glutathione peroxidase
SOD1 is mainly distributed in the cytoplasm, although it was also found in the nucleus, mitochondria and lysosomes (Zelko et al., 2002). To date, over 180 mutations have been identified in all five exons of the SOD1 gene, although not all of these mutations are pathogenic (Tang et al., 2018). Of these mutations, an alanine 4 change to valine (A4V) is the most common in American fALS patients, and an aspartic acid 90 change to alanine (D90A) is the most common in fALS patients worldwide (Anderson, 2006). After almost 30 years, the exact mechanism by which SOD1 mutations lead to ALS pathogenesis still needs to be elucidated, but numerous hypotheses have been proposed, some of which will be discussed below. Most propositions suggest that ALS is a consequence of gain of function of the mutant SOD1 protein, instead of loss of function, since mice lacking the SOD1 gene do not develop ALS (Chen et al., 2013). Mutations generally result in a decrease in protein stability and sequentially, aggregations of SOD1 proteins take shape (Kaur et al., 2016).

2.1 Glutamate excitotoxicity
One of the hypotheses concerning ALS pathogenesis as a result of mutant SOD1, involves glutamate. Glutamate is the major excitatory neurotransmitter in the central nervous system and is synthesized in the presynaptic terminal. Once it has been released, it diffuses across the synaptic cleft to activate postsynaptic receptors which causes Ca^{2+} ions to enter the postsynaptic neuron and an action potential to be triggered. A constant raise of intracellular calcium levels, however, initiates several processes within the neuron that eventually lead to neuronal damage and cell death, referred to as glutamate excitotoxicity. To prevent this, exciting amino acid transporters (EAATs) present on neurons and glia remove glutamate from the synaptic cleft (Sundaram et al., 2012).

Mutant SOD1 transgenic mice have been found to have decreased levels of EAAT2, which would result in an increase of glutamate within the synaptic cleft and thus, stimulate glutamate excitotoxicity and neurodegeneration. Therefore, it is suggested that mutant SOD1 may contribute to motor neuron degeneration by raising glutamate excitotoxicity (Trotti et al., 1999). Glutamate excitotoxicity is probably not the main cause, though, since mutant SOD1 transgenic mice overexpressing EAAT2 did not show longer survival compared to transgenic mice with normal EAAT2 expression (Guo et al., 2003). Furthermore, the before-mentioned riluzole administration for ALS treatment engages on glutamate excitotoxicity by inhibiting glutamate release, but does not cure ALS (Miller et al., 2012).

2.2 Mitochondrial dysfunction
Furthermore, mitochondrial function is also thought to be affected as a result of mutant SOD1, causing ALS. In the spinal cords of ALS patients, and in mutant SOD1 transgenic mice as well, mutant SOD1 aggregates in mitochondria (Higgins et al., 2003), which is believed to interfere with multiple pathways within these organelles. In the presence of mutant SOD1, a reduction in respiration was observed and as a result, less ATP was produced (Mattiazzi et al., 2002). It was also reported that mitochondria in neurons of mutant SOD1 transgenic mice showed less Ca^{2+} buffering capacity, which could result in an increase of cytosolic calcium levels and thus, in heightening the risk for excitotoxic events (Damiano et al., 2006). Other mitochondrial pathways affected by mutant SOD1, involve apoptotic triggering and increased superoxide production. Mitochondria of mutant SOD1 transgenic animals overexpress proapoptotic proteins like Bim and Bax, while expression of anti-apoptotic proteins Bcl-2 and Bcl-Xl has
been found to be decreased (Vukosavic et al., 1999). As a result, programmed cell death is more likely to occur. It was previously mentioned that superoxide free radicals could inflict oxidative damage.

In addition, mutant SOD1 appears to disrupt the transport of mitochondria along the axons. Motor neurons have long axons with high energy demands, making them vulnerable to stress. Consequently, it is of great importance that organelles, RNA, proteins, et cetera are transported to the end of the axons. It has been implied, however, that mutant SOD1 transgenic mice have abnormal retro- and anterograde transport, which would be damaging to neurons (Pasinelli and Brown, 2006). Impaired axonal transport has been hypothesized to be the result of mitochondrial dysfunction and glutamate excitotoxicity (Zarei et al., 2015), both of which have been explained above.

2.3 Influence of microglial and astroglial cells

Another proposed hypothesis is that the expression of mutant SOD1 in microglial and astroglial cells is key to ALS pathogenesis. This theory arose due to the observation that mutant SOD1 transgenic mice expressing the mutant SOD1 only in their motor neurons did not show neuronal degeneration, while transgenic mice with non-mutated motor neurons surrounded by SOD1-mutated glial cells did (Clement et al., 2003). Moreover, by replacing the SOD1-mutated glial cells with healthy, wild-type glia or by downregulating the expression of mutant SOD1 in these cells, motor neuron degeneration significantly decreased and the mutant SOD1 transgenic mice lived longer (Lee et al., 2012).

It is believed that the expression of mutant SOD1 in microglial cells, the resident macrophages within the brain and the first line of defense against infection and injury, promotes the neurotoxic phenotype of microglial cells, thus stimulating disease progression (Bonafede and Mariotti, 2017). The expression of mutant SOD1 in astroglial cells, which normally play an important role in maintaining and supporting neurons, would also result in such a phenotype (Bonafede and Mariotti, 2017).

3 Transactive response DNA-binding protein (TARDBP)

Another major gene linked to ALS is the TARDBP gene (Kabashi et al., 2008; Sreedharan et al., 2008), which is found to be mutated in about 4.2% of European fALS patients and in about 1.5% of Asian fALS patients (Zou et al., 2017). The TARDBP gene is located on chromosome 1p36.22, exists of six exons and encodes for the transactive response DNA-binding protein 43, or TDP-43 (Gendron et al., 2013). This protein, a 43-kDa heterogeneous ribonucleoprotein (hnRNP) with a length of 414 amino acids, is involved in many processes concerning RNA metabolism, including transcription, alternative splicing, translation, and mRNA transport (Hergesheimer et al., 2019). To execute these, TDP-43 is mostly located in the nucleus, although the protein is also present in the cytoplasm (Gendron et al., 2013). TDP-43 was first described to inhibit expression of the HIV-1 gene by binding to chromosomally integrated transactive response DNA (Ou et al., 1995).

Over 40 different TARDBP-mutations have been found in relation to ALS, most of which are located on exon 6 of the gene (Gendron et al., 2013). This region encodes for the highly conserved, glycine-rich, C-terminus of TDP-43, a crucial domain in governing the functional properties of the protein. It controls the interactions of TDP-43 with other proteins and the region is involved in inhibiting the splicing of certain RNA transcripts, as well. The C-terminus also influences the solubility and cellular localization of TDP-43 (Gendron et al., 2013; Hergesheimer et al., 2019). Mutations in this region are
proposed to lead to TDP-3 mislocalization from the nucleus to the cytoplasm and the formation of large cytoplasmic aggregates of TDP-43, as are seen in ALS patients (Ayala et al., 2008).

It has been suggested that due to the nuclear exclusion of TDP-43, the protein can no longer normally function, disrupting the RNA metabolism of its targets and eventually causing neurodegeneration (Hergesheimer et al., 2019). Research of Iguchi et al. supported this theory, since mice lacking the TARDBP gene showed age-dependent progressive motor neuron degeneration (Iguchi et al., 2013). Among the targets of TDP-43, Statmin2 (STMIN2), which is essential to regulating the cytoskeleton, is implied to play a role. Recently, a study making use of human stem cell–derived motor neurons demonstrated that once the levels of TDP-43 were decreased, gene expression of STMIN2 dropped simultaneously with TDP-43. Furthermore, rescuing expression of STMIN2 helped regain motor neuron growth in the cells (Klim et al., 2019). Other studies have implied that neurodegeneration could also be a consequence of the toxic properties of the TDP-43 aggregates, such as sequestering proteins which subsequently become nonfunctional (Hergesheimer et al., 2019).

Interestingly, aggregation of TDP-43 is very common in all ALS cases, also in patients who do not carry a mutation in the TARDP gene. Around 95% of patients with wild type TARDBP manufacture TDP-43-positive aggregates. It is still up for debate what the underlying cause is for this phenomenon (Hergesheimer et al., 2019).

4 Fused in sarcoma (FUS)

Not long after the TARDBP gene, the FUS gene, located on chromosome 16p11.2, was also implicated in ALS pathogenesis (Kwiatkowski et al., 2009). This gene has been found to be mutated in about 2.8% of fALS patients in Europe and in about 6.4% of Asian fALS patients (Zou et al., 2017) and encodes for the protein fused in sarcoma (FUS). The name of this protein finds its origin in the observation that, as a result of a chromosomal translocation, FUS is fused with another protein called CHOP in malignant human myxoid liposarcoma, activating the transcription of oncogenes and promoting tumorigenesis (Rabbitts et al., 1993). The FUS protein is 526 amino acids long, with a mass of 75 kDa, and has a similar function as the TDP-43 protein. It regulates multiple processes within RNA metabolism, including transcription, alternative splicing, and mRNA transport, just like TDP-43. FUS is involved in DNA damage regulation, as well (Gal et al., 2011). In neurons and glial cells, FUS is mostly present in the nucleus (Gal et al., 2011).

The majority of FUS-mutations causing ALS concentrate in the region encoding for the C-terminus of the protein (Kino et al., 2011). This domain contains the nuclear localization sequence, which is responsible for the transport of FUS into the nucleus. Thus, mutations in the C-terminus lead to mislocalization of the FUS protein to the cytoplasm, where it accumulates and will form aggregations, just like TDP-43 (Gal et al., 2011). ALS-associated mutations, however, do not elicit complete mislocalization as some amount of the FUS protein remains present in the nucleus (Kino et al., 2011). It has been suggested that ALS consequently develops as a result of loss of normal function of the FUS protein in the nucleus, but a toxic gain of function in the cytoplasm has been hypothesized to be causative, too (Ishigaki and Sobue, 2018). In support of the loss-of-function hypothesis, mutations in the FUS gene linked to ALS lead to reduced binding of FUS to chromatin, which is necessary for transcription activation (Yang et al., 2014). Decreased splicing activity of the FUS protein has also been reported in
context with mutations within the respective gene (Sun et al., 2015), as well as a deficiency in DNA double strand break repair (Gao et al., 2017). However, it is generally believed that the toxic gain-of-function hypothesis is more likely to be involved with ALS pathogenesis, since mice lacking the FUS gene show behavioral impairments, but no ALS-like symptoms (Kino et al., 2015).

The toxic gain-of-function hypothesis states that accumulated mutant-FUS proteins in the cytoplasm have a neurotoxic effect, leading to motor neuron degeneration (Ishigaki and Sobue, 2018). It is believed that these proteins influence the dynamics of stress granules, which are non-membranous, cytoplasmic ribonucleoprotein (RNP) granules composed of mRNAs, translation initiation factors, ribosomes, and other RNA binding proteins. Stress granules appear in response to cellular stresses that inhibit translation initiation, such as oxidative stress or a viral infection, and once the cell is no longer stressed, the granules usually disassemble (Gao et al., 2017). In the presence of mutant FUS, though, it seems that irreversible assembly formation is more likely, which would mean that all factors inside are trapped. This could affect multiple processes within RNA metabolism. In what manner this would exactly lead to neurodegeneration, and thus to ALS pathogenesis, is still unclear (Murakami et al., 2015).

5 Chromosome 9 open reading frame 72 (C9ORF72)
In Europe and North America, the most frequently mutated gene linked to ALS pathology, is the C9ORF72 gene (DeJesus-Hernandez et al., 2011). It is mutated in about 33.7% of the European fALS patients (Zou et al., 2017) and in about 36% of the North American fALS patients (Majounie et al., 2012). Interestingly, mutations in the C9ORF72 gene are much less common in Asia: only 2.3% of the fALS patients of Asian descent show a mutation within this gene (Zou et al., 2017). C9ORF72 is located on chromosome 9p21.2, in the 72nd open reading frame (ORF) of the chromosome, which explains the gene’s nomenclature.

The C9ORF72 gene consists of 11 exons and due to alternative splicing, the gene has three different transcript variants, which translate to two different protein isoforms (Balendra and Isaacs, 2018). Transcript variant 1 encodes for the short isoform (C9ORF72-S) with a length of 222 amino acids protein and a mass of 24 kDa, while transcript variants 2 and 3 encode for the long isoform (C9ORF72-L), which is 481 amino acids long and with a mass of 54 kDa (Balendra and Isaacs, 2018). The exact function of these protein isoforms is poorly understood, but in recent years, research has made progress in illuminating the role of C9ORF72. Bioinformatic analysis has shown that the C9ORF72 protein is, as regards to secondary structure, similar to differentially expressed in normal and neoplastic cells (DENN)-domain containing proteins, which operate as guanine nucleotide exchange factor (GEF) proteins (Levine et al., 2013). These proteins activate GTPases of the Rab family by stimulating guanosine diphosphate (GDP) release and therefore allow binding of guanosine triphosphate (GTP) (Marat et al., 2011). The similarity between GEF-proteins and C9ORF72 would imply that the C9ORF72 protein has a comparable function, which recently has been confirmed (Iyer et al., 2018). By activating proteins of the Rab family, vesicle trafficking is enabled, as well as macroautophagic activity (Marat et al., 2011). The Rab proteins mediate membrane movements from the autophagosomes, leading to the engulfment and degradation of cytoplasmic contents (Nassif et al., 2017). Since Rab proteins are involved in vesicle trafficking and macroautophagy, and it seems they are activated by C9ORF72, this would suggest C9ORF72 is connected to vesicle trafficking and macroautophagy, too (Nassif et al., 2017).
Indeed, depletion of C9ORF72 gene expression using short hairpin RNA (shRNA) and short interfering RNA (siRNA) partially impaired autophagic activity in mouse primary cortical neurons and human cell lines, respectively. Subsequently, accumulation and aggregation of proteins were observed (Sellier et al., 2016; Webster et al., 2016). Interestingly, only expression of C9ORF72-L was able to rescue autophagy dysfunction, while C9ORF72-S did not. Such an observation implies that the role of the short isoform of C9ORF72 is unrelated to autophagy (Balendra and Isaacs, 2018). Conversely, overexpression of the C9ORF72 gene increased the amount of autophagosomes in human cell lines, indicating a rise in autophagic activity (Webster et al., 2016). Other research has indicated that silencing the C9ORF72 gene also inhibits endocytosis, a form of vesicle trafficking, in human cell lines, since the transport of a certain endocytosis marker from the plasma membrane to the Golgi apparatus was reduced (Farg et al., 2014). Impaired endosomal and lysosomal trafficking as a consequence of C9ORF72 depletion has been found, as well (Balendra and Isaacs, 2018). All findings support the idea that the C9ORF72 protein functions in macroautophagy and vesicle trafficking.

The mutation that is found in ALS patients, is a significant hexanucleotide GGGGCC (G\(_4\)C\(_2\)) repeat expansion in the first intron between exons 1a and 1b of C9ORF72. Neurologically healthy people tend to have 2 to 24 hexanucleotide repeats, while hundreds to thousands of repeats are common in patients with ALS (Nguyen et al., 2018). It is thought that smaller repeat expansion sizes correlate with later ALS onset (Nguyen et al., 2018). How the hexanucleotide repeat expansion leads to ALS, is unknown, but three major hypotheses have been proposed, concerning loss of function, RNA toxicity and proteotoxicity from dipeptide repeat (DPR) aggregates (Nguyen et al., 2018).

5.1 Loss of function
Induced pluripotent stem cell (iPSC)-derived neurons, the frontal cortex, the cerebellum, the motor cortex and the cervical spinal cord acquired from ALS patients with mutated C9ORF72 all have been found to have reduced levels of C9ORF72 transcripts (Balendra and Isaacs, 2018). These observations have led to believe that it is potentially the loss of C9ORF72 protein that causes ALS. A decrease in C9ORF72 transcripts would lead to lower C9ORF72 protein production and consequently, there would be a reduction in autophagic activity and vesicle trafficking. Since these processes are essential to cell survival, reduced activity could hypothetically result in degeneration of the motor neurons and thus, in ALS (Balendra and Isaacs, 2018).

This hypothesis, however, does not fully explain the development of ALS, because it has been demonstrated that mice without the C9ORF72 gene, do not show motor neuron degeneration, muscle weakness or other ALS-related symptoms (Koppers et al., 2015). It is worth mentioning that mice lacking the C9ORF72 were found to have abnormal immune cell function, which would suggest that it is possible that immune cells have an impact on the development of ALS in C9ORF72-mutated patients (Sudria-Lopez et al., 2016).

5.2 RNA toxicity
Bidirectional transcription of the hexanucleotide repeat expansion in the C9ORF72 gene is known to form repetitive sense (GGGGCC) and antisense (CCCCGG) RNA transcripts, which have been reported to accumulate in the neuronal nuclei of ALS patients with mutated C9ORF72 (Balendra and Isaacs, 2018). These RNA transcripts shape into RNA foci, previously found in diseases such as myotonic dystrophy types 1 and 2 and fragile X tremor ataxia syndrome. In these diseases, the RNA foci sequester essential
RNA-binding proteins and thus, impair normal RNA metabolism, resulting in pathogenesis (La Spada et al., 2010).

A similar mechanism has been proposed to cause ALS (Balendra and Isaacs, 2018). In support of this theory, a series of RNA binding proteins has been shown to bind and colocalize with RNA foci in tissue of ALS patients with mutated C9ORF72 and it has been revealed that the presence of the hexanucleotide repeat expansion dysregulated splicing (Youn-Bok et al., 2013). Moreover, injecting embryos of zebrafish, a model for ALS, with repetitive sense and antisense RNA transcripts reduced axon length and lead to abnormal axon branching, implying motor axonopathy (Swinnen et al., 2018). However, there is evidence challenging the idea that ALS pathology is consequential to RNA toxicity. Research in the Drosophila model proved no neurotoxic effect of the RNA foci. Although RNA foci accumulated in neuronal nuclei and withdrew RNA binding proteins, neurodegeneration did not occur (Moens et al., 2018).

5.3 Proteotoxicity from dipeptide repeat (DPR) aggregates
Since recent years, it is known that repetitive RNA motifs can be translated, even though an AUG start codon is absent. Through such repeat-associated non-AUG (RAN) translation, the repetitive sense and antisense RNA transcripts of the C9ORF72 gene are translated to several dipeptide repeats (DPRs). A total of five different DPRs can be produced in relation to the mutated C9ORF72, due to the presence of six reading frames in either the sense and antisense direction (Figure 1) (Balendra and Isaacs, 2018). From these DPRs, poly-GA (glycine-alanine) is the most abundantly produced, followed by poly-GP (glycine-proline), poly-GR (glycine-arginine), poly-PA (proline-alanine) and poly-PR (proline-arginine) (Freibaum and Taylor, 2017). In several brain regions of ALS patients with mutated C9ORF72, including the motor cortex, the DPRs have been found to form aggregates, mostly in the neuronal cytoplasm. In lower numbers, the aggregates are also present in the spinal cord (Freibaum and Taylor, 2017). It is believed that the DPRs disrupt multiple cellular mechanisms and therefore lead to ALS pathogenesis (Balendra and Isaacs, 2018).

![Figure 1. Repeat-associated non-AUG (RAN) translation of the hexanucleotide repeat expansion in the C9ORF72 gene. The dipeptide repeats (DPRs) poly-GR, poly-GP, poly-GA, poly-PR and poly-PA are produced in the six possible open reading frames in sense and antisense direction.](image)

Corresponding to this hypothesis, a great variety of studies have observed malfunction of cellular mechanisms in the presence of DPRs. For example, DPRs affected mitochondrial function and alternative splicing and induced endoplasmic reticulum stress. Most importantly, DPRs have been shown
to cause neurodegeneration (Balendra and Isaacs, 2018). The arginine-containing DPRs, poly-GR and poly-PR, appear to be most toxic, but poly-PA and poly-GP are unlikely to be toxic species (Balendra and Isaacs, 2018).

Contradictory to the proposition that it are the DPRs that cause ALS, no correlation has been found between the amount of DPRs present in a brain region and the severity of neurodegeneration in that region. The greatest amount of DPRs were localized in the cerebellum, while neurodegeneration was greatest in the motor cortex and spinal cord (Mackenzie et al., 2013). This counter argument has partly been refuted, though, since the researchers of the before-mentioned studies looked at post-mortem brains of ALS patients that were at the end-stage of their disease and used immunohistochemistry which could only detect large aggregates of DPRs (Freibaum and Taylor, 2017). Perhaps there is a correlation between DPR burden and severity of neurodegeneration in patients at earlier disease stages. Moreover, by only detecting large aggregates of DPRs soluble DPRs and smaller aggregates are overlooked, resulting in an underestimation of the amount of DPRs present in a brain region (Freibaum and Taylor, 2017).

### 6 Other genes linked to ALS

Besides the SOD1 gene, the TARDBP gene, the FUS gene and the C9ORF72 gene, more than 25 additional genes have been associated with ALS pathogenesis, with a significant rise in the amount since recent years due to progress in technological advances in large scale genetic screening (Table 1) (Nguyen et al., 2018). These genes, however, account for a much smaller population of fALS patients in which they are mutated compared to the before-mentioned genes. Some genes have only been found to be mutated within a few families, such as the UBQLN2 gene (Deng et al., 2011) and CHMP2B gene (Parkinson et al., 2006), and some are uniquely connected to very rare variants of ALS, such as the ALS2 gene (Yang et al., 2001; Hadano et al., 2001) and SPG11 gene (Orlacchio et al., 2010), both of which have been linked to juvenile-onset ALS. This variant of ALS is characterized by a much earlier onset of upper and lower motor neuron degeneration compared to the classical variant of ALS. Patients show impairment of voluntary movements prior to the age of 25, but it is fortunate that progression is significantly slower (Swinnen and Robberecht, 2014).

Despite low mutation frequencies, these genes have given important insight into the mechanisms behind motor neuron degeneration as seen in ALS, just like SOD1, TARDBP, FUS and C9ORF72 have done. For instance, the PFN1 gene, which is mutated in less than 1% of fALS patients and encodes for a small actin-binding protein that promotes formin-based actin polymerization, was the first gene to implicate the involvement of cytoskeletal dysfunction of neurons in familial ALS pathogenesis (Wu et al., 2012). By looking at shared functional characteristics of the protein products of mentioned genes, specific pathways in ALS pathogenesis have been elucidated, such as RNA metabolism, vesicle trafficking, autophagy and cytoskeletal maintenance. Engagement in these pathways could perhaps help with the development of new therapeutic strategies (Boylan, 2015).

Almost all genes associated with ALS and referenced in Table 1, have been found to be located on an autosomal chromosome, with chromosome 2 and 12 being the most concentrated with ALS-linked genes. Only the UBQLN2 gene is linked to the X chromosome (Deng et al., 2011). Most of these genes are dominantly inherited, except for the ALS2 gene, SPG11 gene, and SIGMAR1 gene, which are
exclusively recessively inherited (Yang et al., 2001; Hadano et al., 2001; Orlacchio et al., 2010; Al-Saif et al., 2011). Mutations in the OPTN gene have been observed to be inherited in both a dominant as well as a recessive manner, though (Maruyama et al., 2010). Interestingly, dependent on which gene is mutated a different ALS phenotype can be found. For example, in the presence of mutated VAPB, patients had a longer disease course than usual (Nishimura et al., 2004).

Table 1: Additional genes linked to ALS pathogenesis

<table>
<thead>
<tr>
<th>Year of discovery*</th>
<th>Gene</th>
<th>Chromosome</th>
<th>Protein</th>
<th>Presumed function**</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001 (Yang et al.; Hadano et al.)</td>
<td>ALS2</td>
<td>2q33.1</td>
<td>alsin</td>
<td>Vesicle trafficking</td>
</tr>
<tr>
<td>2002 (Hand et al.)</td>
<td>ALS3</td>
<td>18q21</td>
<td>unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>2003 (Sapp et al.)</td>
<td>ALS7</td>
<td>20p13</td>
<td>unknown</td>
<td>Unknown</td>
</tr>
<tr>
<td>2004 (Chen et al.)</td>
<td>SETX (ALS4)</td>
<td>9q34.13</td>
<td>senataxin</td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>2004 (Nishimura et al.)</td>
<td>VAPB (ALS8)</td>
<td>20q13.32</td>
<td>vesicle-associated membrane protein B</td>
<td>Vesicle trafficking</td>
</tr>
<tr>
<td>2006 (Greenway et al.)</td>
<td>ANG (ALS9)</td>
<td>14q11.2</td>
<td>angiogenin</td>
<td>Hypoxia-induced angiogenesis</td>
</tr>
<tr>
<td>2006 (Parkinson et al.)</td>
<td>CHMP2B (ALS17)</td>
<td>3p11.2</td>
<td>charged multivesicular body protein 2B</td>
<td>Vesicle trafficking</td>
</tr>
<tr>
<td>2009 (Chow et al.)</td>
<td>FIG4 (ALS11)</td>
<td>6q21</td>
<td>phosphoinositide 5-phosphatase</td>
<td>Vesicle trafficking</td>
</tr>
<tr>
<td>2010 (Orlacchio et al.)</td>
<td>SPG11 (ALS5)</td>
<td>15q15–21.1</td>
<td>spatacsin</td>
<td>Vesicle trafficking</td>
</tr>
<tr>
<td>2010 (Maruyama et al.)</td>
<td>OPTN (ALS12)</td>
<td>10p13</td>
<td>optineurin</td>
<td>Influencing NF-κB</td>
</tr>
<tr>
<td>2010 (Elden et al.)</td>
<td>ATXN2 (ALS13)</td>
<td>12q24.12</td>
<td>ataxin-2</td>
<td>Golgi maintenance Vesicle trafficking</td>
</tr>
<tr>
<td>2010 (Johnson et al.)</td>
<td>VCP (ALS14)</td>
<td>9p13.3</td>
<td>valosin-containing protein</td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>2011 (Deng et al.)</td>
<td>UBQLN2 (ALS15)</td>
<td>Xp11.21</td>
<td>ubiquilin-2</td>
<td>Autophagy</td>
</tr>
<tr>
<td>2011 (Al-Saif et al.)</td>
<td>SIGMAR1 (ALS16)</td>
<td>9p13.3</td>
<td>sigma non-opioid intracellular receptor 1</td>
<td>Neuroprotection</td>
</tr>
<tr>
<td>2011 (Fecto et al.)</td>
<td>SQSTM1 (FTD-ALS3)</td>
<td>5q35.3</td>
<td>sequestosome-1</td>
<td>Autophagy</td>
</tr>
<tr>
<td>2012 (Wu et al.)</td>
<td>PFN1 (ALS18)</td>
<td>17p13.2</td>
<td>profilin-1</td>
<td>Cytoskeletal maintenance</td>
</tr>
<tr>
<td>2013 (Takahashi et al.)</td>
<td>ERBB4 (ALS19)</td>
<td>2q34</td>
<td>receptor tyrosine-protein kinase erbB-4</td>
<td>Mitogenesis Differentiation</td>
</tr>
<tr>
<td>2013 (Kim et al.)</td>
<td>HNRNPA1 (ALS20)</td>
<td>12q13</td>
<td>heterogeneous nuclear ribonucleoprotein A1</td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>2013 (Kim et al.)</td>
<td>HNRNPA2B1</td>
<td>7p15.2</td>
<td>heterogeneous nuclear ribonucleoprotein A2B1</td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>2014 (Johnson et al.)</td>
<td>MATR4 (ALS21)</td>
<td>5q31.2</td>
<td>matrin-3</td>
<td>RNA metabolism</td>
</tr>
<tr>
<td>Year</td>
<td>Gene</td>
<td>Chromosome</td>
<td>Function</td>
<td>Reference</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>2014 (Smith et al.)</td>
<td><strong>TUBA4A</strong> (ALS22)</td>
<td>2q35</td>
<td>tubulin alpha-4A</td>
<td>Cytoskeletal maintenance</td>
</tr>
<tr>
<td>2014 (Bannwarth et al.)</td>
<td><strong>CHCHD10</strong> (FTD-ALS2)</td>
<td>22q11.23</td>
<td>Coiled-coil-helix-coiled-helix domain containing protein 10</td>
<td>Mitochondrial function</td>
</tr>
<tr>
<td>2015 (Freischmidt et al.; Pottier et al.)</td>
<td><strong>TBK1</strong> (FTD-ALS4)</td>
<td>12q14.2</td>
<td>TANK-binding kinase 1</td>
<td>Inflammation, Autophagy</td>
</tr>
<tr>
<td>2016 (Brenner et al.)</td>
<td><strong>NEK1</strong> (ALS24)</td>
<td>4q33</td>
<td>NIMA related kinase 1</td>
<td>DNA damage repair</td>
</tr>
<tr>
<td>2017 (Smith et al.)</td>
<td><strong>ANXA11</strong> (ALS23)</td>
<td>10q22.3</td>
<td>annexin A11</td>
<td>Vesicle trafficking</td>
</tr>
<tr>
<td>2018 (Nicolas et al.; Brenner et al.)</td>
<td><strong>KIF5A</strong> (ALS25)</td>
<td>12q13.3</td>
<td>kinesin family member 5A</td>
<td>Cytoskeletal maintenance</td>
</tr>
</tbody>
</table>

* Year in which the first mutation(s) was or were found within the mentioned gene in fALS patients, except for the ALS3 and ALS7 genes. For these genes, the year represents when the region was linked to ALS using linkage analysis. No mutations have been reported yet for both ALS3 and ALS7.

** Function as described in the original research papers that had found the gene to be mutated.

### 7 Discussion and conclusion

To date, over 30 different genes have been identified to be involved in ALS pathogenesis and the list of additional genes with mutations in familial ALS patients continues to grow, aided by technological advances in large scale genetic screening (Nguyen et al., 2018). Of these genes, the *SOD1* gene, *TARDBP* gene, *FUS* gene and *C9ORF72* gene have been found to be most frequently mutated in European fALS patients, with mutation frequencies of 14.8%, 4.2%, 2.8% and 33.7% respectively (Zou et al., 2017). Interestingly, the mutation frequencies for these and the other ALS-linked genes differ per geographic region. For example, mutations within the *C9ORF72* gene are the most common cause of fALS in Europe and North America (Zou et al., 2017; Majounie et al., 2012), while this gene is mutated in only 2.3% of the fALS patients of Asian descent, making it only the fourth most common genetic cause of ALS in this region (Zou et al., 2017). It has been speculated that the origin of this mutation dates back about 1500 years to a Finnish population and that the hexanucleotide repeat expansion, as is seen in mutated *C9ORF72*, was spread through Europe by Vikings and their descendants (Pliner et al., 2014). About 70% of the European fALS patients is explained by the genes that are currently associated with ALS (Renton et al., 2014).

Most of the ALS-causative genes have been observed to produce proteins that, under healthy, physiological conditions, function within processes such as RNA metabolism, vesicle trafficking, autophagy and cytoskeletal maintenance. How mutations within these genes result in ALS pathogenesis, though, is largely debated. Some mutated genes are thought to produce proteins which no longer have the ability to execute their tasks, leading to dysfunctional RNA metabolism, for instance, with neuronal cell death as a result. This has been hypothesized to be one of the pathologic mechanisms for mutated *TARDBP* (Hergesheimer et al., 2019). Other mutated genes are believed to yield alternative forms of their corresponding proteins that have gained a new toxic function, inducing motor neuron degeneration. Mutations within the *SOD1* gene have been suggested to lead to glutamate excitotoxicity,
mitochondrial dysfunction, and consequently, to impaired axonal transport within the neuronal cells (Trotti et al., 1999; Zarei et al., 2015). Mutant SOD1 present is also believed to cause a neurotoxic phenotype in microglial and astroglial cells, which would stimulate disease progression (Bonafede and Mariotti, 2017). A gain of function has been implied to be involved with mutations in the FUS gene, as well. It is thought that mutant-FUS proteins sequester mRNAs, translation initiation factors, ribosomes, and other RNA binding proteins, resulting in disruption of RNA metabolism (Gao et al., 2017; Murakami et al., 2015). Mutations within the C9ORF72 gene are proposed to have both a loss of function and a toxic gain of function. In the absence of wild type C9orf72, autophagic activity and vesicle trafficking are observed to be reduced (Balendra and Isaacs, 2018), while in the presence of mutant C9orf72, RNA foci and DPRs are produced, which are seen as toxic. RNA foci are suspected to sequester essential RNA-binding proteins, similarly to mutant-FUS proteins, and DPRs are thought to impair multiple cellular mechanisms, such as mitochondrial function and alternative splicing, eventually causing ALS pathogenesis (Balendra and Isaacs, 2018). In what manner the function changes of the genes would exactly lead to ALS pathogenesis, still needs to be clarified, but there is an abundance of proposed pathways on which therapies could engage (Figure 2).

![Figure 2](image)

**Figure 2. The proposed pathological mechanisms behind mutations within the SOD1 gene, the TARDBP gene, the FUS gene and the C9ORF72 gene, resulting in ALS.**

Similarly, the reason why motor neurons are affected by these mutations, and for example sensory neurons are not, has not yet been explained. The SOD1 gene, TARDBP gene, FUS gene and C9ORF72 gene are expressed all over the body (GeneCards), which would mean that mutant forms of the corresponding proteins should be present within all cells, if these genes are mutated. Nevertheless, only the upper and lower motor neurons degenerate in ALS. It is proposed that the large size of these neurons and their broad cytoskeleton are what make them more vulnerable for dysfunctional cellular mechanisms (Shaw and Eggett, 2000). Due to their size, the upper and lower motor neurons request
more metabolic support compared to other cells and, consequently, optimal mitochondrial function and vesicle trafficking are necessary to fulfill the demand (Shaw and Eggett, 2000. Furthermore, motor neurons have been found to be highly sensitive to excitotoxicity and dysregulation of intracellular calcium homeostasis, which could contribute to the vulnerability of the motor neurons to disruptions within these pathways (Shaw and Eggett, 2000). The consensus is that at least two genes causative for ALS have to be mutated to result in upper and lower motor neuron degeneration (Van Blitterswijk et al., 2012).

Although the before-mentioned genes have given some insight into ALS pathogenesis of fALS patients, the majority of ALS patients does not belong to this category. About 90% of ALS patients is ‘sporadic’, as they do not have first or second degree relatives suffering from ALS (Bonafede and Mariotti, 2017). For these individuals, it is believed that certain environmental risk factors, such as exposure to heavy metals, electromagnetic fields and pesticides induce oxidative stress, and in combination with the presence of specific susceptibility genes cause ALS (Bonafede and Mariotti, 2017; Riancho et al., 2018). This raises the question of whether the mutated genes and the corresponding pathological mechanisms found in fALS patients, are relevant to the sALS patients and their perspective on new therapies. It appears that the mechanisms do just that for at least a select portion of the sALS patients. About 7.4% of European sALS patients have been found to carry a mutation within the SOD1, TARDBP, FUS and C9ORF72 genes, just like fALS patients (Zou et al., 2017), which would suggest that these individuals might actually belong to the fALS category. Perhaps they were miscategorized as a result of a small family size, incomplete family history, or incorrect diagnoses within the family, leading to the absence of first and second degree relatives with ALS (Baylon, 2015). This would imply that categorization between fALS and sALS patients is inconvenient. Moreover, it suggests that genetic screening should be more available for all ALS patients, even if they appear to have no obvious genetic link. At present, genetic screening is usually only done if a mutation within a gene is suspected (Baylon, 2015), but by including it in the standard protocol for all ALS patient, patients could be informed on their genetic load and the identification of ALS-susceptibility genes could be helped, as well.

In conclusion, due to technical advances in the last decades, the current knowledge of genetics behind ALS pathogenesis has become extensive, but more research is crucial to understand the main causative mechanisms behind ALS so new effective therapy strategies can be developed. The pathological pathways of the mutated genes are mostly hypothetical and a significant portion of the fALS patients remains unexplained by a mutated gene (Renton et al., 2014). Furthermore, research should focus more on the genetic screening of sporadic ALS patients, which entail the majority of patients with ALS, to achieve a higher understanding of ALS pathogenesis.

8 References


