Decoding ALS: from genes to mechanism

J. Paul Taylor1, Robert H. Brown Jr1 & Don W. Cleveland4,5

Amyotrophic lateral sclerosis (ALS) is a progressive and uniformly fatal neurodegenerative disease. A plethora of genetic factors have been identified that drive the degeneration of motor neurons in ALS, increase susceptibility to the disease or influence the rate of its progression. Emerging themes include dysfunction in RNA metabolism and protein homeostasis, with specific defects in nucleocytoplasmic trafficking, the induction of stress at the endoplasmic reticulum and impaired dynamics of ribonucleoprotein bodies such as RNA granules that assemble through liquid–liquid phase separation. Extraordinary progress in understanding the biology of ALS provides new reasons for optimism that meaningful therapies will be identified.

Amyotrophic lateral sclerosis (ALS) is a prototypical neurodegenerative disease that is characterized by the progressive degeneration of motor neurons in the brain and spinal cord. The condition was first described by the neurologist Jean-Martin Charcot, and its name reflects both the degeneration of corticospinal motor neurons, the descending axons of which in the lateral spinal cord seem scarred (lateral sclerosis), and the demise of spinal motor neurons, with secondary denervation and muscle wasting (amyotrophy). Corticospinal neurons make direct or indirect connections with spinal motor neurons, which innervate skeletal muscles and trigger their contraction (Fig. 1a). This Review summarizes the clinical and pathological features of ALS and describes how discoveries in ALS genetics have illuminated important themes in the molecular pathophysiology of the disease.

ALS is known as Lou Gehrig's disease in the United States and as motor neuron disease in the United Kingdom. Although onset of the disease occurs commonly in mid-adulthood (at a mean age of 55 years), ALS might begin as early as in the first or second decade of life or could even emerge in later life. Similar to most neurodegenerative diseases, it starts focally and spreads: symptoms that start as subtle cramping or weakness in the limbs or bulbar muscles progress to the paralysis of almost all skeletal muscles. Some subsets of motor neurons, including those that innervate the extraocular muscles or sphincters, are spared until late in the progression of the disease. However, ALS is invariably fatal. Death occurs typically 3–5 years after diagnosis, although some forms of the disease demonstrate protracted survival.

ALS is an orphan disease that is diagnosed in 1–2 individuals per 100,000 each year in most countries; the prevalence of ALS is about 5 cases per 100,000 people, which reflects the rapid lethality of the disease1. In the United States and United Kingdom, ALS causes more than 1 in 500 deaths in adults, a statistic that suggests that more than 15 million people who are alive at present will succumb to the disease.

The clinical manifestations of ALS

Considerable heterogeneity exists in the general rubric of ALS. Clinical subsets of the disease are distinguished by the involvement of different sets of motor neurons or different regions of the body. Depending on the location of the main pathology, those affected might develop weakness with flaccidity and atrophy of the limbs (known as progressive muscular atrophy, which mainly affects spinal neurons or lower motor neurons), prominent hyperreflexia and spasticity with increased limb tone but little muscle atrophy (known as primary lateral sclerosis, which affects corticospinal motor neurons with limited involvement of spinal motor neurons), tongue atrophy with thickness of speech and difficulty swallowing (known as bulbar ALS, which affects brainstem motor neurons that serve the muscles of tongue movement, chewing, swallowing and articulation) or slow and highly dysfunctional speech and swallowing in the absence of tongue atrophy, often accompanied by the accentuation of emotional reflexes (known as pseudobulbar palsy, which affects cortical frontal bulbar motor neurons). Importantly, ALS shares clinical and pathological features with several other adult-onset degenerative disorders, including, most frequently, frontotemporal dementia (FTD), which could constitute a clinical spectrum (Box 1).

Genetic contributions to ALS

About 10% of ALS cases are transmitted in families, almost always as a dominant trait and frequently with high penetrance. The first genetic mutations found to cause ALS, reported in 1993, affected the gene SOD1 (ref. 2) and more than 50 further potential ALS genes have been published since, although validating the causality of specific variants remains a challenge. By applying rigorous criteria, a list of genes with mutations that are implicated unequivocally in the pathogenesis of ALS can be generated (Table 1). These genes can be grouped into several loose categories: genes that alter proteostasis and protein quality control; genes that perturb aspects of RNA stability, function and metabolism; and genes that disturb cytoskeletal dynamics in the motor neuron axon and distal terminal. The mutations involved are mostly missense substitutions, although the genetic lesion in C9orf72 is an enormous expansion of an intronic hexanucleotide repeat.

Although sporadic ALS should refer strictly to disease that presents without a family history of ALS, this term is sometimes mistakenly used to refer to ALS that occurs without a genetic basis. Technological advances that facilitate broad DNA sequencing in people with sporadic ALS have revealed that genetic variants in established ALS genes are not infrequent. For example, it is now evident that 1–3% of sporadic cases of ALS are caused by missense mutations in SOD1 (ref. 3) and another 5% or more are caused by intronic expansions in C9orf72 (ref. 4). Pathogenic mutations in other ALS genes, including TARDBP, FUS, HNRNPA1, SQSTM1, VCP, OPTN and PFN1, have also been identified.

1Howard Hughes Medical Institute and the Department of Cell and Molecular Biology, St. Jude Children’s Research Hospital, Memphis, Tennessee 38105, USA. 2Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01655, USA. 3Ludwig Institute for Cancer Research, University of California, San Diego, La Jolla, California 92093, USA. 4Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA.

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.
in people with sporadic ALS, although they are rare.

Genetic variants that enhance susceptibility to ALS or that modify the clinical phenotype are of immense interest, even if the variants themselves do not cause ALS. For example, large expansions of repeats of the trinucleotide CAG in the coding sequence of the gene ATXN2 cause spinocerebellar ataxia type 2, in which motor weakness is sometimes an early presentation. It is striking therefore that modest expansions to an early presentation. It is striking therefore that modest expansions to

spinal motor neurons shrink and accumulate rounded or thread-like deposits of aggregated proteins that are referred to collectively as inclusions (Fig. 1b). The cytoplasmic inclusions in ALS often become ubiquitinated; an initial target for ubiquitination is TAR DNA-binding protein 43 (TDP-43), encoded by the gene TARDBP, which forms the main component of ubiquitinated inclusions in most cases of ALS.

Other pathological features are associated with specific genes. For example, cases of ALS caused by a large expansion of a hexanucleotide repeat in C9orf72 show intranuclear RNA foci, as well as neuronal cytoplasmic inclusions, predominantly in the cerebellum and hippocampus, that contain the protein sequestosome-1 (also known as ubiquitin–binding protein p62 and encoded by the gene SQSTM1) but are distinct from TDP-43 inclusions that are also present in such individuals (ref. 9). Cases of ALS caused by mutations in the genes SOD1 or FUS are pathologically distinct because they exhibit inclusions of abnormal SOD1 or FUS proteins, respectively, rather than those of TDP-43. In addition to these findings in motor neurons, there is also abundant evidence of relevant pathology in non-neural cell types (for example, insidious astroglialosis and microgliosis). It is probable that both of these forms of non-cell-autonomous cellular reactivity influence adversely the progression of ALS.

**Pathogenic mechanisms of ALS**

The molecular era of discovery in ALS began with the identification of dominant mutations in the gene SOD1, which encodes an abundant, ubiquitously expressed cytoplasmic enzyme called Cu–Zn superoxide dismutase. An important antioxidant, the normal function of SOD1 is to catalyse the conversion of highly reactive superoxide (most frequently produced by errors in mitochondria) to hydrogen peroxide or oxygen.

The expression of mutant SOD1 in mice demonstrated that the degeneration of motor neurons is driven by one or more acquired toxicities of the mutant protein and is independent of dismutase activity. The more than 170 ALS-causing mutations that have now been identified (http://alsod.iop.kcl.ac.uk/) lie in almost every region of the 153-amino-acid SOD1 polypeptide. Moreover, although many variants retain partial or full dismutase activity, there is no correlation between a reduction in activity and the age of disease onset or the speed of disease progression. These findings led to the consensus that disease arises from one or more toxic properties of the many SOD1 mutants rather than from reduced dismutase activity.

A sobering reality, however, is that in the 23 years since the discovery of mutations in SOD1, no consensus on the main toxicity of mutant SOD1 has emerged. Instead, a plethora of toxic mechanisms that mediate the degeneration and death of motor neurons have been proposed (Fig. 2). A prominent finding is that a proportion of each ALS-causing SOD1 mutant fails to fold properly, which implicates the accumulation of misfolded SOD1 as a possible contributor to toxicity in ALS. Misfolded SOD1 forms ubiquitinated cytoplasmic inclusions that can occur early in ALS and that escalate as the disease progresses.

The accrual of ubiquitinated SOD1 aggregates in people with SOD1 mutations is paralleled by the accrual of ubiquitinated TDP-43 aggregates in people with TARDBP mutations (as well as in most people with sporadic ALS), which highlights a correlation between protein aggregation and ALS. However, as has been demonstrated for other neurodegenerative diseases, large aggregates of disease-causing mutant SOD1 are not sufficient to drive disease because their elimination fails to affect any aspect of the fatal disease that develops in mice expressing ALS-linked mutants of SOD1 (ref. 15).

**Non-cell–autonomous toxicity**

Similar to the genes implicated in other main neurodegenerative diseases, all genes in which ALS-causing mutations occur are expressed in many cell types. Indeed, it is now clear that ALS arises, in part, through non-cell-autonomous mechanisms. This means that the disease is the result of a combination of damage from mutant SOD1 in both motor neurons and their glial partners, rather than from damage to neurons alone.

For mutant SOD1, this concept is underscored by studies in mice

---

**Figure 1 | Components of the nervous system that are affected by ALS.** a. ALS mainly affects the descending corticospinal motor neurons (upper motor neurons) that project from the motor cortex into synapses in the brainstem and spinal cord, and the bulb or spinal motor neurons (lower motor neurons) that project into skeletal muscles. b. Subtypes of ALS show typical pathological features: SOD1 aggregates (arrows) in spinal motor neurons in SOD1-related familial ALS (top left); TDP-43 redistribution to cytoplasmic inclusions (arrows) in spinal motor neurons in sporadic ALS (top right); RNA foci in the nucleus (arrows) and the cytoplasm (arrowhead) of a cortical neuron affected by C9 ALS–FTD (bottom left); GA (bottom centre) and GR (bottom right) dipeptide-repeat pathology in the dentate nucleus of a brain affected by C9 ALS–FTD (bottom right).
revealing that high levels of mutant SOD1 expression in all motor neurons is not sufficient to cause early onset disease16. Conversely, a reduction in the synthesis of mutant SOD1 in motor neurons does not slow the rate of progression after the onset of disease, even when applied before symptoms occur17–19. Therefore, ALS is a disease not just of the motor neuron but also of the motor system, which is comprised of motor neurons and intimately associated cells of several types.

The crucial role of glia in ALS

The importance of glial cells in the degeneration and death of motor neurons emerged from studies in which the synthesis of mutant SOD1 was silenced in microglia, astrocytes or oligodendrocyte precursor cells. Microglia, which are the innate immune cells of the nervous system, become activated in all types of ALS (Fig. 2b). The synthesis of mutant SOD1 by microglia is an important determinant of rapid disease progression, as determined by selectively silencing the mutant gene SOD1 in microglia19 or by using cell grafts to replace microglia expressing mutant SOD1 with normal microglia20. Consistent with these findings, inhibition of the transcription factor NF-kB suppresses this neuroinflammatory component of microglial toxicity in co-cultured motor neurons21.

A further mechanism of damage that results from mutant SOD1 produced by microglia is counterintuitive: stimulation of the excessive extra-cellular production of superoxide22. Misfolded mutant SOD1 can associate with the small GTPase RAC1, which controls the activation of NADPH oxidase, a complex that produces superoxide (Fig. 2b). So, instead of its normal function of removing intracellular superoxide, mutant SOD1 could drive microglia to produce high levels of extracellular superoxide.

Disturbances in microglial function have also emerged as a potential contributor to ALS that is associated with mutations in C9orf72. Recognition that mutations in C9orf72 result in the decreased expression of C9orf72 in people with ALS8 suggests that the loss of C9orf72 function might contribute to disease. The protein that C9orf72 encodes is a potential guanine exchange factor for one or more as-yet-unidentified G proteins. Its inactivation in mice results in abnormal microglia and age-related neuroinflammation, providing evidence that non-cell-autonomous, microglial-mediated inflammation might contribute to ALS23–25.

A crucial contribution of mutant SOD1 to pathogenesis is driven by oligodendrocytes, which are cells that myelinate the axons of upper motor neurons and the initial axonal segments of lower motor neurons. A reduction in the synthesis of mutant SOD1 early in oligodendrocyte maturation produces a more striking delay in the onset of disease26 than does similar suppression of mutant SOD1 synthesis in motor neurons16,19. Oligodendrocytes also support motor neuron function by directly supplying the energy metabolite lactate to the axon through the action of monocarboxylate transporter 1 (MCT 1) (Fig. 2c). Mutant SOD1 impairs the expression of MCT 1 by oligodendrocytes in mouse models of ALS27. A similar reduction in the accumulation of MCT 1 is found in sporadic ALS27, which is consistent with a non-cell-autonomous role for the reduced supply of energy from oligodendrocytes as a general component of ALS pathogenesis.

Another type of glial cell, the astrocyte, provides motor neurons with nutrients, ion buffering and recycling of the neurotransmitter glutamate. The selective reduction of mutant SOD1 synthesis by astrocytes in mice slowed the onset28 or progression19 of disease. This delay was accompanied by a delay in the activation of microglia, demonstrating a functional crosstalk between mutant-SOD1-expressing astrocytes and microglia.

One of the earliest proposed mechanisms to underlie ALS was glutamate excitotoxicity, which is the excessive firing of motor neurons that is derived from a failure to rapidly remove synaptic glutamate (Fig. 2d). Astrocytes limit the firing of motor neurons through the swift recovery of glutamate, a function that is mediated by excitatory amino acid transporter 2 (EAAT2), which transports glutamate into the astrocyte (Fig. 2d). The loss of EAAT2 has been observed both in SOD1-mutant rodent models of ALS14,19 and in samples from people with familial or sporadic ALS30. The resulting failure of astrocytes to quickly clear synaptic glutamate triggers the repetitive firing of action potentials and a corresponding increase in calcium influx, as well as endoplasmic reticulum (ER) and mitochondrial stress as the result of overwhelming the calcium storage capacities of these organelles.

Astrocytes also protect motor neurons from excitotoxic damage through the release of an unidentified soluble factor or factors that induce motor neurons to upregulate the glutamate receptor subunit GluR-2 (ref. 31). The incorporation of GluR-2 into glutamate receptors in neurons reduces the permeability of these receptors to calcium, which provides protection from excitotoxicity by decreasing the influx of calcium. Astrocytes that express mutant SOD1 fail to regulate GluR-2 expression in co-cultured neurons, thereby increasing their vulnerability to excitotoxic damage31.

Several teams of researchers have used in vitro co-cultures of motor neurons and astrocytes (or astrocyte-conditioned medium) to show that astrocytes expressing ALS-linked mutations produce a toxicity that diffuses to motor neurons32–36 (Fig. 2d). However, there is no consensus on the identity of the toxic species. Notably, astrocytes from people with familial ALS or sporadic ALS (obtained directly from autopsy samples36 or by isolating neuronal precursor cells that can be converted into astrocytic precursor cells and then astrocytes36) are toxic to co-cultured normal motor neurons. This finding36 is especially provocative because it indicates that neuronal precursor cells in portions of tissue from people with sporadic ALS have already acquired damage and that this damage is retained following several divisions of these cells in culture and their subsequent differentiation. Whether toxicity from sporadic ALS-derived astrocytes is mediated by changes in SOD1 (ref. 35) or not36 remains unsettled.

Most importantly, a non-cell-autonomous contribution of astrocytes to ALS-like disease has been demonstrated in rodents expressing mutant SOD1 in which transplantation to the spinal cord of lineage-restricted astrocyte precursors without SOD1 mutations delayed progression of the disease27.

ALS genes induce ER stress or impair protein degradation

ER stress has been implicated broadly in ALS (Fig. 2a). Initial evidence arose from studies of mutant SOD1 in which misfolded SOD1 binds to the cytoplasmic surface of the ER integral membrane protein derlin-1 (ref. 38). This binding leads to the inhibition of ER-associated degradation (ERAD), the pathway for extraction and degradation of misfolded proteins from the ER. Moreover, relieving ER stress delays the progression of disease in an animal model of ALS26.

There is now overwhelming evidence to show that disruption of the
two main protein clearance pathways, the ubiquitin–proteasome system and autophagy, can be central components of the disease mechanism in ALS (Fig. 2a). Several ALS-causing mutations occur in genes with products that are involved directly in protein degradation, including ubiquitin-2 (ref. 40) and sequestosome-1 (ref. 41), both of which function as adapters that bring polyubiquitinated proteins to the proteasome or the autophagosome for degradation. Mutations have also been reported in optineurin (OPTN), a proposed receptor for autophagy (43), and valosin-containing protein (VCP) (44), which has a role in ERAD and sorting endosomal proteins. Other studies have shown FTD-linked mutations in CHMP2B (45), which encodes a protein that has been implicated in maturation of the autophagosome and endosomal cargo sorting and degradation. ALS-linked mutations are also found in VAPB (46), the product of which functions in the unfolded protein response in the delivery of ER-ejected substrates to the proteasome. A preponderance of biochemical evidence has demonstrated a decrease in the activities of the proteasome in lumbar spinal cords before symptoms occur in mice that express mutant SOD1 (ref. 47) or following the sustained expression of mutant SOD1 in a cultured line of neurons (48).

Axon disorganization and disrupted transport in ALS

Disorganization of the axonal cytoskeleton, and especially of the neurofilaments, is a conspicuous feature both of familial ALS and sporadic ALS (Fig. 2e). As the most asymmetric cells in nature, and with axons that can reach more than 1 metre in length, motor neurons must rely on axonal transport to deliver components that are synthesized in the cell bodies to axons and synapses. ALS-linked mutant SOD1 has been demonstrated to slow both anterograde (49) and retrograde (46, 51) transport routes months before neurodegeneration. Indeed, reduction in retrograde transport through mutations in dynactin (52), which is an activator of the retrograde motor cytoskeletal dynein (53), provokes motor neuron disease in humans.

Owing to the peculiar architecture of neurons, it is a challenge for these cells to alter local gene expression at the synapse in response to neuronal input or changes in the synaptic environment. To achieve this, neurons must transport all necessary components for translation (for example, messenger RNA, ribosomes and translation factors) to distal sites for local protein synthesis (54). The spatial distribution of mRNAs depends on the proper microtubule-dependent transport of neuronal RNA transport granules and other factors, and it is regulated by several RNA-binding proteins that are associated with ALS, including TDP-43, FUS and heterogeneous nuclear ribonucleoprotein (hnRNPA) A1. ALS-causing mutations in TDP-43 impair the axonal transport of RNA granules in Drosophila and in cultured neurons, including motor neurons derived from people with ALS (55).

### A prion–like spread in inherited ALS

The prion-like, templated conversion of a natively folded protein into a misfolded version of itself is now recognized as a prominent feature of the cell-to-cell spread of protein aggregates in neurodegenerative diseases. Examples include α-synuclein templating in Parkinson’s disease, amyloid-β aggregation in Alzheimer’s disease and tau misfolding in chronic brain injury (reviewed in further detail in ref. 56). Evidence for similar templated toxicity has emerged for misfolded SOD1 (refs 57 and 58), with wild-type SOD1 exacerbating the toxicity of mutant SOD1 in mice (59). Prion-like propagation and development of disease is initiated focally by the injection of lysates containing mutant SOD1 into mice that express mutant SOD1 (ref. 60). This finding replicates the correlation between focal initiation and spreading in people with familial ALS or sporadic ALS (61).

That said, prion-like propagation of SOD1 (or other ALS-linked proteins) has not been achieved in rodents without the coexistence of a pre-existing, weakly active mutant ALS gene. It is unresolved whether this evidence challenges the prion-like spread model of sporadic ALS or, alternatively, whether it raises the possibility that there must be a pre-existing sensitivity in individuals who develop sporadic ALS that facilitates such spreading. Coupled with the recognition that misfolded mutant SOD1 can be secreted by motor neurons or astrocytes (62), potentially through the newly discovered pathway in which misfolded proteins are secreted unconventionally as an adaptation to proteasome dysfunction (63), stochastic focal initiation provides a plausible mechanism for the age-dependent onset of disease and its subsequent spread. As most cases of ALS are marked by aggregated TDP-43 rather than SOD1, an unresolved question is whether TDP-43 also exhibits templated misfolding that can spread from cell to cell.

### Table 1 | The genetics of ALS

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene</th>
<th>Protein function</th>
<th>Protein function</th>
<th>Mutations</th>
<th>Proportion of ALS</th>
<th>Date of discovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>21q22.1</td>
<td>SOD1</td>
<td>Cu–Zn superoxide dismutase</td>
<td>Superoxide dismutase</td>
<td>&gt;150</td>
<td>20%</td>
<td>1993 (ref. 2)</td>
</tr>
<tr>
<td>2p13</td>
<td>DCTN1</td>
<td>Dynactin subunit 1</td>
<td>Component of dynein motor complex</td>
<td>10</td>
<td>1%</td>
<td>2003 (ref. 52)</td>
</tr>
<tr>
<td>14q11</td>
<td>ANG</td>
<td>Angiogenin</td>
<td>Ribonuclease</td>
<td>&gt;10</td>
<td>&lt;1%</td>
<td>2006 (ref. 141)</td>
</tr>
<tr>
<td>q36</td>
<td>TARDBP</td>
<td>TDP-3</td>
<td>RNA-binding protein</td>
<td>&gt;40</td>
<td>5%</td>
<td>2008 (refs 67 and 142)</td>
</tr>
<tr>
<td>16p11.2</td>
<td>FUS</td>
<td>FUS</td>
<td>RNA-binding protein</td>
<td>&gt;40</td>
<td>5%</td>
<td>2009 (refs 68 and 69)</td>
</tr>
<tr>
<td>9p13.3</td>
<td>VCP</td>
<td>Ubiquitin segregase</td>
<td>Ubiquitin segregase</td>
<td>5</td>
<td>1–2%</td>
<td>2010 (ref. 44)</td>
</tr>
<tr>
<td>10p15-14</td>
<td>OPTN</td>
<td>Optineurin</td>
<td>Autophagy adaptor</td>
<td>1</td>
<td>4%</td>
<td>2010 (ref. 42)</td>
</tr>
<tr>
<td>9p21-22</td>
<td>C9orf72</td>
<td>C9orf72</td>
<td>Possible guanine nucleotide exchange factor</td>
<td>Intronic GGGGCC repeat</td>
<td>25%</td>
<td>10%</td>
</tr>
<tr>
<td>Xp11.23-Xp13.1</td>
<td>UBQLN2</td>
<td>Ubiquitin 2</td>
<td>Autophagy adaptor</td>
<td>5</td>
<td>&lt;1%</td>
<td>2011 (ref. 40)</td>
</tr>
<tr>
<td>5q35</td>
<td>SQSTM1</td>
<td>Sequestosome 1</td>
<td>Autophagy adaptor</td>
<td>10</td>
<td>&lt;1%</td>
<td>2011 (refs 41 and 143)</td>
</tr>
<tr>
<td>17p13.2</td>
<td>PPFN1</td>
<td>Profilin-1</td>
<td>Actin-binding protein</td>
<td>5</td>
<td>&lt;1%</td>
<td>2012 (ref. 144)</td>
</tr>
<tr>
<td>12q13.1</td>
<td>HNRNPA1</td>
<td>hnRNP A1</td>
<td>RNA-binding protein</td>
<td>3</td>
<td>&lt;1%</td>
<td>2013 (refs 70 and 71)</td>
</tr>
<tr>
<td>5q31.2</td>
<td>MATR3</td>
<td>Matrin 3</td>
<td>RNA-binding protein</td>
<td>4</td>
<td>&lt;1%</td>
<td>2014 (ref. 76)</td>
</tr>
<tr>
<td>2q36.1</td>
<td>TUBA4A</td>
<td>Tubulin α-4A chain</td>
<td>Microtubule subunit</td>
<td>7</td>
<td>&lt;1%</td>
<td>2014 (ref. 145)</td>
</tr>
<tr>
<td>22q11.23</td>
<td>CHCHD10</td>
<td>Coiled-coil-helix-coiled-coil-helix domain-containing protein 10</td>
<td>Mitochondrial protein of unknown function</td>
<td>2</td>
<td>&lt;1%</td>
<td>2014 (ref. 146)</td>
</tr>
<tr>
<td>12q14.1</td>
<td>TBK1</td>
<td>Serine/threonine-protein kinase TBK1</td>
<td>Regulates autophagy and inflammation</td>
<td>10</td>
<td>?</td>
<td>2015 (ref. 147)</td>
</tr>
</tbody>
</table>
The intersection of RNA biology and ALS pathogenesis

In 2006, Virginia Lee and colleagues reported the mislocalization of RNA-binding protein TDP-43 from its predominantly nuclear location to ubiquitin-containing cytoplasmic inclusions in affected areas of the brain and the spinal cord of people with ALS. TDP-43 mislocalization is now recognized widely as the hallmark of both sporadic ALS and most familial forms of ALS. This seminal discovery has implications beyond ALS because TDP-43 mislocalization to cytoplasmic inclusions is also the hallmark of FTD that lacks tau-containing inclusions (about half of all cases of FTD) and inclusion body myopathy (IBM), which are diseases that show genetic overlap with ALS. Moreover, TDP-43 pathology is also found as a secondary pathological feature in a subset of people with Alzheimer’s disease or Parkinson’s disease. The importance of TDP-43 in pathogenesis was cemented by the identification of ALS-causing mutations in this protein. The subsequent identification of ALS-causing mutations in related proteins that bind RNA, including FUS and hnRNP A1, focused substantial attention on the role of RNA biology in ALS pathogenesis.

TDP-43, FUS and hnRNP A1 are members of the hnRNP family of proteins that regulate RNA metabolism at every stage of the RNA life cycle. They bind to thousands of RNA targets, which implies that a disturbance in the function of one or more of these proteins has the potential to affect RNA metabolism on a broad scale. Further links between the pathogenesis of ALS and RNA biology have emerged from the identification of ALS-causing mutations in the RNA-binding protein matrin-3 (ref. 76), an appreciation of the increased risk of developing ALS in association with certain alleles of the RNA-binding protein matrin-3 (ref. 77) and the recognition of RNA-related mechanisms of disease that are associated with mutations in C9orf72 (refs 8 and 77).

Phase separation gives rise to membraneless organelles

RNA metabolism occurs in complex RNA–protein assemblies that can coalesce into a variety of membraneless organelles such as nucleoli and stress granules. Interestingly, these organelles behave as complex liquids that arise through phase separation, a process in which protein-laden RNAs separate from the surrounding aqueous nucleoplasm or cytoplasm in a manner that is akin to the separation of oil from vinegar. Phase separation is mediated by low-complexity domains that are present in RNA-binding proteins such as TDP-43, FUS and hnRNP A1 (refs 79–81). The assembly of membraneless organelles is a strategy of cellular compartmentalization that governs many biological processes. However, the contribution of phase transition to this process presents a risk because RNA-binding proteins with low-complexity domains, which are prone to fibrillization, are placed in close proximity. Indeed, mutations that cause ALS are found frequently in the low-complexity domains of TDP-43 (ref. 82), FUS and hnRNP A1 (ref. 70). As a consequence, these mutations alter the dynamics of membraneless organelles and also accelerate fibrillization, which results in the formation of amyloid-like fibrils that are deposited in the cell bodies and the neuropil (Fig. 3c).

Mutations in the low-complexity domains of at least six different hnRNP result in a clinicopathological spectrum that ranges from ALS and FTD to IBM (Fig. 3a and Box 1). Notably, some disease-causing mutations in RNA-binding proteins do not affect low-complexity domains. For example, several ALS-causing mutations in FUS and hnRNP A1 disturb the nuclear localization sequence of these proteins and result in their accumulation in the cytoplasm. Phase transition by RNA-binding proteins that contain low-complexity domains is exquisitely dependent on concentration, and it is probable that the increased accumulation of FUS and hnRNP A1 in the cytoplasm as a consequence of mutations that affect the nuclear localization signals of these proteins is sufficient to drive excess phase separation, as shown by the hyperassembly of stress granules in cells derived from people with relevant mutations (refs 79,80).

RNA metabolism defects in ALS

Disturbance of the normal phase transitions carried out by RNA-binding proteins can have deleterious consequences, including altering the material properties of RNA granules and impairing their function. Moreover, persistent assembly of RNA-binding proteins in the highly concentrated liquid state may promote the formation of amyloid-like fibrils that have toxic properties and may also result in a partial or complete loss of the normal function of important RNA-binding proteins. A well known feature of ALS histopathology is the redistribution of TDP-43 from the neuronal cell body and distal compartments.

Figure 2 | Mechanisms of disease implicated in ALS. a, Familial ALS-associated mutations frequently affect genes that are components of the cellular protein quality control system. Other mutations, such as those in SOD1, affect protein folding. b, Hyperactivation of microglia produces extracellular superoxide, which triggers inflammation and degeneration in motor neurons. c, A reduction in the levels of the lactate transporter MCT1 diminishes energy supplied by oligodendrocytes to motor neurons. d, A failure of astrocytes to clear synaptic glutamate via the transporter EAAT2 triggers repetitive firing of motor neurons and excitotoxicity. e, Disruption of the cytoskeleton and impaired axonal transport limits the exchange of essential macromolecules and organelles between the neuronal cell body and distal compartments. f, Disturbances in aspects of RNA metabolism, including RNA processing, transport and utilization, are largely the result of impaired hnRNP function.
nucleus to the cytoplasm. A similar redistribution is observed for FUS and hnRNPA1 when disease-causing mutations occur in the genes that encode these proteins. This redistribution of proteins might reflect a cytoplasmic sink that is produced by the hyperassembly of cytoplasmic granules or by poorly dynamic RNA granules that fail to disassemble appropriately, the deposition of amyloid-like fibrils and defects in nucleocytoplasmic trafficking, as well as other potential mechanisms.

The culmination of this redistribution is the depletion of RNA-binding proteins in the nucleus that has the potential to cause a considerable loss of nuclear function. A well-known function of TDP-43 in the nucleus is the regulation of alternative splicing. Experimental depletion of TDP-43 in rodents was found to alter hundreds of splicing events in the brain, resulting in the depletion of several RNAs that encode synaptic proteins. The loss of nuclear TDP-43 also facilitates the use of cryptic splice sites that, in general, might lower the levels of correctly spliced protein-encoding mRNAs. Furthermore, TDP-43 autoregulates its synthesis, which establishes the possibility of a feed-forward mechanism that amplifies the impact of the partial loss of TDP-43 function.

The loss of FUS or hnRNPA1 from the adult nervous system produces defects analogous to those associated with the loss of TDP-43, although different subsets of mRNAs are linked to the depletion of each of these RNA-binding proteins. An important, unanswered question concerns the extent to which ALS caused by other genetic perturbations, especially C9orf72-related ALS, also involves disturbances in RNA biology that intersect mechanistically with mutations in TDP-43, FUS and hnRNPA1.

The biogenesis of regulatory RNA and its function in ALS

Both TDP-43 and FUS are components of macromolecular complexes that generate small non-coding RNAs known as microRNAs (miRNAs) with functions in RNA silencing. The loss of TDP-43 or FUS results in a reduction in the expression of miRNAs in model systems, including Drosophila models and induced pluripotent stem (iPS)-derived motor neurons from people with TDP-43 mutations, which suggests a possible role for altered RNA silencing in ALS. Various miRNAs contribute to the regulation of neuromuscular junctions, implying that motor neurons might be particularly sensitive to disturbances in the biogenesis of miRNA. Indeed, global downregulation of miRNAs has been reported in motor neurons from people with sporadic ALS, although the role of reduced levels of miRNAs in the pathogenesis of ALS remains to be established. Nonetheless, the expression of regulatory RNAs seems to be altered robustly and consistently in the serum of people with ALS, and this could present an opportunity for the development of biomarkers.

The curious case of C9orf72–related ALS and FTD

Although the identification of ALS-causing mutations that affect SOD1 and RNA-binding proteins highlighted pathophysiological pathways through which disease might arise, most of the genetic burden of ALS remained unaccounted for until 2011. Genetic linkage studies followed by several large genome-wide association studies identified the location of a gene in the chromosome 9p21 locus in which mutations cause both ALS and FTD. During sequencing of the non-coding regions of candidate genes in chromosome 9p21, a pathogenic expansion of a hexanucleotide repeat in C9orf72 was identified as the basis for C9orf72–related ALS and FTD. In healthy individuals, the sequence GGGGCC was present as 2–23 repeats but in affected individuals it was expanded to hundreds or thousands of repeats. In parallel, an independent study also discovered a pathogenic expansion of GGGGCC repeats in C9orf72 (ref. 77).

The consequences of repeat expansion in C9orf72

Three non-exclusive mechanisms have been proposed through which expanded GGGGCC repeats might cause C9 ALS–FTD (Fig. 4). First, a reduction in the expression levels of C9orf72 was identified in the brains of mice and in people with C9 ALS–FTD. Global downregulation of miRNAs has been reported in motor neurons from people with sporadic ALS, although the role of reduced levels of miRNAs in the pathogenesis of ALS remains to be established. Nonetheless, the expression of regulatory RNAs seems to be altered robustly and consistently in the serum of people with ALS, and this could present an opportunity for the development of biomarkers.

Three non-exclusive mechanisms have been proposed through which expanded GGGGCC repeats might cause C9 ALS–FTD (Fig. 4). First, a reduction in the expression levels of C9orf72 was identified in the brains of mice and in people with C9 ALS–FTD. Global downregulation of miRNAs has been reported in motor neurons from people with sporadic ALS, although the role of reduced levels of miRNAs in the pathogenesis of ALS remains to be established. Nonetheless, the expression of regulatory RNAs seems to be altered robustly and consistently in the serum of people with ALS, and this could present an opportunity for the development of biomarkers.
and lymphadenopathy. These mice were found to have abnormal macrophages and microglia, as well as age-related neuroinflammation and signs of autoimmunity, which raises the possibility of a non-cell-autonomous inflammatory contribution to C9 ALS–FTD.

Nevertheless, the dominant inheritance pattern of C9 ALS–FTD, the absence of people with ALS or FTD with null alleles or missense mutations in C9orf72 and the absence of neurodegeneration in C9orf72-knockout mice provide arguments against the loss of C9orf72 function as the sole driver of disease. Indeed, most empirical evidence points to the gain of toxic functions as the main mechanisms that drive neurodegeneration in C9 ALS–FTD. For example, the adeno-associated virus-mediated delivery to the brain of a construct that expresses expanded GGGGCC repeats elicits neurodegeneration, although the nature of the toxic species in C9 ALS–FTD remains unclear.

**Gain of toxic function from repeat–containing RNA**

The initial description of the mutation in C9orf72 was accompanied by evidence to show that RNA foci containing the GGGGCC repeat accumulate in the brains and spinal cords of people with C9 ALS–FTD (Fig. 1b), and this suggested a second possible disease mechanism, involving toxic gain of function by repeat-containing RNA (Fig. 4). It was then noted that the gene C9orf72 can be transcribed bidirectionally and that foci containing sense (GGGGCC) or antisense (CCCCGG) RNA transcripts accumulate in affected cells. The accrual of such foci in C9 ALS–FTD is reminiscent of the pathological RNA foci that are observed in myotonic dystrophy type 1, myotonic dystrophy type 2 and fragile X-associated tremor and ataxia syndrome, which are also caused by the expansion of nucleotide repeats in non-coding regions. In these diseases, the accumulated repeat-containing RNA sequesters RNA-binding proteins that are involved in splicing, which leads to defects in splicing that underlie some aspects of pathogenesis. Similarly, a number of RNA-binding proteins bind to expanded GGGGCC or GGGCCC repeats in vitro, and a rare co-localization with RNA foci has been observed for several of these proteins in tissue from affected individuals.

Simple model systems have illustrated the functional consequences of the sequestration of some hexanucleotide repeat-binding proteins, including transcriptional activator protein Pur-a and Ran GTPase-activating protein 1 (RanGAP1), but the contribution of these interactions to the development of disease is not yet established. Notably, repeats of GGGGCC (but not of CCCCCG) can adopt a stable secondary structure known as a G-quadruplex, which might contribute to the persistence of this species of RNA as well as to its ability to reach distal neurites and associate with RNA-binding proteins in transport granules and potentially interfere with local translation.

**Gain of toxic function from dipeptide repeats**

Substantial evidence has also accrued to implicate a third disease mechanism in C9 ALS–FTD: specifically, toxicity from DPR proteins that are produced by repeat-associated non-AUG (RAN) translation (Fig. 4). This unconventional type of translation occurs in the absence of an initiating AUG codon and might rely on secondary structures formed by repeat-expanded RNA. In C9 ALS–FTD, RAN translation occurs in all reading frames and from both sense and antisense transcripts, and it results in the production of five DPR proteins: glycine-alanine (GA) and glycine-arginine (GR) from sense GGGGCC transcripts; proline-arginine (PR) and proline-alanine (PA) from antisense GGGCCG transcripts; and glycine-proline (GP) from both sense and antisense transcripts. All of these DPR proteins are produced in people with C9 ALS–FTD and they associate for the neuronal cytoplasmic and intranuclear inclusions that contain ubiquitin and sequestosome-1 but lack TDP-43 that are found widely in the brain and spinal cord. The timing, location and level of expression of each species of DPR protein in the brains of affected people are yet to be clarified. Several reports have described the deposition of DPR proteins in the brains of people with C9 ALS–FTD, and in some instances an inverse relationship has been described between the regional burden of DPR proteins and the corresponding severity of neurodegeneration. These studies were based on the post-mortem examination of brains with end-stage disease and relied on the detection of large inclusions using immunohistochemistry, an approach that probably under-represents the pathological burden of soluble DPR proteins. However, the apparent discrepancy between the burden of DPR-protein deposition, the levels of which are greatest in the cerebellum, and the severity of neurodegeneration, which is greatest in the motor cortex and spinal cord, needs to be resolved to understand the role of DPR proteins in the development of disease.

Some species of DPR proteins have been shown to be toxic in cultured cells and animal models of disease, although high levels of expression were sometimes used to produce short-term toxicity. The arginine-containing DPR proteins GR and PR seem to be most toxic. For example, when GR or PR is added to cells in culture, it enters and accumulates in nucleoli, which leads to defects in RNA processing and subsequent cell death. Similarly, the independent expression of each of the five species of DPR proteins in cultured neurons revealed that GR and PR are very toxic, whereas PA, GA and GP are well tolerated. Observations in Drosophila engineered to express each of the five DPR proteins have also shown that GR and PR are extremely toxic to neuronal tissue, whereas GA is modestly toxic and GP and PA seem to be non-toxic. A recent discovery is that the arginine-containing DPR proteins GR and PR bind to proteins that contain low-complexity domains. Furthermore, GR and PR alter the phase separation of such proteins, resulting in the perturbed assembly, dynamics and function of membraneless organelles such as stress granules and nucleoli. This finding mirrors the defects in phase transitions that are observed with disease-causing mutations in the low-complexity domains of TDP-43, FUS and hnRNP1, suggesting a common pathological mechanism.

However, other investigations have reported that toxicity is associated with the expression of GA in cell culture and its adeno-associated virus-mediated delivery to the mouse brain. It should be noted that these efforts to model the toxicity of DPR proteins have used short (fewer than 100) repeats. How the properties of those short DPR proteins compare with the possibly larger products of RAN translation in affected individuals is also unknown.

**A defect in nucleocytoplasmic trafficking**

Whereas the nature of the gain of toxic function is still an open question,
progressing evidence suggests that impaired nucleocytoplasmic trafficking is one of the downstream consequences of mutations in C9orf72. A comprehensive, unbiased screen in Drosophila for genetic modifiers of the toxicity that is mediated by expanded GGGGCC repeats identified 18 genes that are connected to the nuclear pore complex and nucleocytoplasmic trafficking\(^{136}\). A separate unbiased screen for genetic modifiers of PR toxicity in yeast also identified numerous genes that encode components of the nuclear pore complex and effectors of nucleocytoplasmic trafficking\(^{136}\). A third study focused on the nucleocytoplasmic transport factor RanGAP1, which binds to the RNA sequence GGGGCC. Genes encoding RanGAP1 and other nucleocytoplasmic transport factors were identified as modifiers of toxicity mediated by expanded GGGGCC repeats in Drosophila\(^{136}\). Consistent with these results, morphological abnormalities were found in the nuclear envelope architecture in both cell-based and animal models of disease, as well as in the brains of people with C9 ALS–FTD. Moreover, defects in the nucleocytoplasmic transport of RNA and proteins were found in neurons derived from the iPSC cells of people with C9 ALS–FTD\(^{136,138,134}\).

### Approaches to therapy for C9 ALS–FTD

The relative contributions of the various proposed modes of toxicity to the development of C9 ALS–FTD is an important consideration that will influence strategies for therapeutic intervention. Efforts are underway to impede Ran translation with small molecules but the success of such an approach will depend on the role of DPR proteins in disease. Irrespective of the main basis for the toxic gain of function, the mutant gene C9orf72 presents an attractive target for therapeutic intervention. For example, antisense oligonucleotides are able to reverse pathological features in neurons derived from iPSC cells\(^{133,135,136}\) or in fibroblasts\(^{136}\) from people with C9 ALS–FTD. Indeed, neurons and glial cells derived from iPSC cells might prove to be a useful model system in which to develop approaches for mitigating toxicity related to mutant C9orf72 even before the basis of toxicity has been elucidated fully.

Therapeutic efforts will be aided further by the development of transgenic mouse models that express human C9orf72 that contain about 450 hexanucleotide repeats, which recapitulate aspects of the molecular pathology\(^{134,137,139}\) neuropsychological deficits\(^{138,139}\) and the motor phenotype\(^{138}\) of C9 ALS–FTD. It is also particularly promising that pathological abnormalities can be reversed, and that the development of neuropsychological deficits can be delayed, by a single-dose infusion of an antisense oligonucleotide that induces the catalytic degradation of hexanucleotide-containing RNAs without exacerbating a reduction in RNAs encoding the C9orf72 protein\(^{133}\).

### Looking forward

Clearly, there has been dramatic progress towards defining the genetic topography and molecular biology of ALS. There is also little doubt that the pace of discovery will continue or even accelerate in several areas of research.

First, it is certain that our understanding of the genetic basis of ALS will continue to evolve. Research programmes are already in place to collect and sequence thousands of whole genomes from people with ALS. More genes that are implicated in ALS are likely to be defined, both through conventional Mendelian genetics and through enhanced association studies that identify increased burdens of rare genetic variants, including those found in non-coding DNA. In parallel, enhanced scoring and recording of quantifiable clinical parameters will permit the definition of variants that modify the phenotype of ALS. The existence of extensive ALS genome databases will enable the first comprehensive studies of epistasis, characterizing the interactions of numerous genes to perturb the viability of motor neurons.

Second, although the past two decades have witnessed extraordinary progress in understanding familial ALS, it is probable that insights that help to elucidate sporadic ALS will be acquired. One view is that all cases of sporadic ALS will ultimately be shown to reflect several genetic determinants. Alternatively, there is increasing interest in exogenous factors that might trigger sporadic neurodegeneration, and atypical infections or the activation of endogenous retroviruses\(^{140}\) are proposed to have such a role. Although the role of external environmental factors in ALS has been elusive, there is fresh interest in the influence of the intrinsic environment, represented by the microbiome, on development of the disease.

Last, and perhaps most importantly, there will be considerable achievements in the development of therapies for ALS. Although daunting, the complexity of the molecular pathology of ALS is promising as a roadmap for defining therapeutic targets. Moreover, for types of ALS that arise from well-defined genetic defects, advances in gene silencing and gene editing technologies will permit personalized therapeutic programmes. When combined with improved methods for the delivery of therapies to the central nervous system, these approaches will lead to strategies for attenuating the lethal course of ALS.


\(3.\) The first report of a gene defect that causes ALS.


\(7.\) The first report of a relationship between microsatellite expansion in ATXN2 and susceptibility to ALS.


\(16.\) This report showed that the expression of mutant SOD1 in microglia accelerates the progression of ALS, establishing a role for non-cell autonomous events in motor neuron degeneration in the disease.


25. The report documented the amelioration of cognitive deficits in mice that express expanded hexanucleotide repeats following the intraventricular infusion of antisense oligonucleotides that reduce the levels of GGGGCC RNA transcripts and DPR proteins.


45. Johnson, J. O. et al. Exome sequencing reveals a set of mutations in TDP43-positive intraneuronal inclusions in a patient with familial ALS. Ann. Neurol. 69, 1451–1460 (2011). This paper and ref. 71 were the first to identify mutations in the gene TDP43 in familial ALS, and also the first demonstration that disease-causing mutations in TDP43 cause familial ALS.


47. Kim, H. J. et al. Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause familial ALS. Cell 147, 647–673 (2011). The first report of a germline mutation in HNRNPA1 as a cause of familial ALS, and also the first demonstration that disease-causing mutations in low-complexity sequences can drive the hyperassembly of a membraneless organelle.


52. Moioli, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrilization. Cell 163, 123–133 (2015). This study demonstrated that low-complexity domains under phase separation that promotes the assembly of membraneless organelles such as stress granules and drives pathological fibrilization.


Acknowledgements
We apologize for the many studies we were unable to cite because of space limitations. The authors gratefully acknowledge the artwork and contributors: N. Kwan and the provision of histopathology images by J. Ravits and S. Saberi. J.P.T. receives support from the Howard Hughes Medical Institute, the US National Institute of Neurological Disorders and Stroke (NINDS), the American Lebanese Syrian Associated Charities, Target ALS and the US ALS Association. R.H.B. receives support from NINDS, the ALS Association, ALS Finding a Cure, ALS ONE, the Angel Fund for ALS Research and Project ALS. D.W.C. receives salary support from the Ludwig Institute for Cancer Research. N. Nedelsky and H.-J. Kim provided editorial assistance. N. Nedelsky and H.-J. Kim and the provision of financial interests: see go.nature.com/2xewlxwy. Readers are welcome to comment on the online version of this paper at go.nature.com/2xewlxwy. Correspondence should be addressed to J.P.T. (peul.taylor@st Jude.org), R.H.B. (robert.brown@umassmed.edu) or D.W.C. (dcleland@ucsd.edu).

Reviewer Information
Nature thanks I. Dikic, J. Rothstein and the other anonymous reviewer(s) for their contribution to the peer review of this work.