Journal of Neurochemistry

JOURNAL OF NEUROCHEMISTRY | 2017 | 141 | 12–30 doi: 10.1111/jnc.13945

REVIEW

Linking amyotrophic lateral sclerosis and spinal muscular atrophy through RNA-transcriptome homeostasis: a genomics perspective

Margarida Gama-Carvalho,* Marina L. Garcia-Vaquero,* Francisco R. Pinto,* Florence Besse,† Joachim Weis,‡ Aaron Voigt,§,¶ Jörg B. Schulz§,¶ and Javier De Las Rivas**

*Universidade de Lisboa, Faculdade de Ciências, BioISI – Biosystems & Integrative Sciences Institute, Campo Grande, 1749-016 Lisboa, Portugal

† Université Cote d'Azur, CNRS, Inserm, iBV, Nice, France

‡Institute of Neuropathology, RWTH Aachen University, Aachen, Germany

§Department of Neurology, University Hospital, RWTH Aachen University, Aachen, Germany

¶JARA-Institute Molecular Neuroscience and Neuroimaging, Forschungszentrum Jülich GmbH and RWTH Aachen University, Aachen, Germany

**Cancer Research Center (CiC-IBMCC, CSIC/USAL/IBSAL), Consejo Superior de Investigaciones Científicas (CSIC) and Universidad de Salamanca (USAL), Salamanca, Spain

Abstract

In this review, we present our most recent understanding of key biomolecular processes that underlie two motor neuron degenerative disorders, amyotrophic lateral sclerosis, and spinal muscular atrophy. We focus on the role of four multifunctional proteins involved in RNA metabolism (TDP-43, FUS, SMN, and Senataxin) that play a causal role in these diseases. Recent results have led to a novel scenario of intricate connections between these four proteins, bringing transcriptome homeostasis into the spotlight as a common theme in motor neuron degeneration. We review reported functional and physical interactions between these four proteins, highlighting their common association with nuclear bodies and small nuclear ribonucleoprotein particle biogenesis and function. We discuss how these interactions are turning out to be particularly relevant for the control of transcription and chromatin homeostasis, including the recent identification of an association between SMN and Senataxin required to ensure the resolution of DNA-RNA hybrid formation and proper termination by RNA polymerase II. These connections strongly support the existence of common pathways underlying the spinal muscular atrophy and amyotrophic lateral sclerosis phenotype. We also discuss the potential of genome-wide expression profiling, in particular RNA sequencing derived data, to contribute to unravelling the underlying mechanisms. We provide a review of publicly available datasets that have addressed both diseases using these approaches, and highlight the value of investing in cross-disease studies to promote our understanding of the pathways leading to neurodegeneration.

Keywords: ALS, RNA metabolism, SMA, transcriptome homeostasis, transcriptomics.

J. Neurochem. (2017) **141**, 12–30.

Motor neuron diseases: causal connections to four key proteins

Motor neuron diseases (MND) are a genetically and clinically heterogeneous group of diseases involving the degeneration of primary motor neurons, leading to muscular weakness and atrophy. Among the genetic causes underlying these and other neurodegenerative diseases, mutations in

Received October 11, 2016; revised manuscript received December 2, 2016; accepted December 24, 2016.

Address correspondence and reprint requests to Margarida Gama-Carvalho, Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, University of Lisbon, Campo Grande, Lisboa, Portugal. E-mail: mhcarvalho@fc.ul.pt

Abbreviations used: ALS, amyotrophic lateral sclerosis; FTD, frontotemporal dementia; MND, motor neuron diseases; SMA, spinal muscular atrophy; SMN, survival of motor neuron.

proteins related to RNA metabolism are relatively frequent (Cooper *et al.* 2009). Examples of such 'RNA-related' disorders are spinal muscular atrophy (SMA) and amyotrophic lateral sclerosis (ALS).

ALS, the most frequent MND, is mostly an adult onset disease with a predominance of sporadic (90%) over hereditary cases, and a relatively high population incidence of 1/20 000 individuals (Chen *et al.* 2013). Pathologically and clinically, ALS is defined by the degeneration of upper and lower motor neurons resulting in muscle weakness and atrophy (symptoms of lower motor neuron dysfunction), but also stiffness, spasticity, hyperreflexia, and pseudobulbar symptoms (upper motor neuron dysfunction). In 30–50% of ALS patients, cognitive symptoms are also observed upon neuropsychological testing.

Familial cases of ALS have been linked to mutations in at least a dozen genes, with a predominance of mutations occurring in *C9ORF72* (linked to 40% of all familial cases), *SOD1*, *FUS,* and *TARDBP* (Chen *et al.* 2013). About 10% of the sporadic cases have been linked to mutations in the same genes, but the remainder are of unknown origin (Renton *et al.* 2014). In almost all cases (with the exception of familial *SOD1* and *FUS* mutations), the cellular pathology shows cytoplasmic aggregation of the TDP-43 protein, encoded by the *TARDBP* gene (Mackenzie *et al.* 2007). These aggregates are characterized by abnormal modifications of TDP-43, including N-terminal truncation, hyperphosphorylation, and ubiquitination. Similar neuropathological characteristics also occur in some cases of frontotemporal dementia (FTD) (Mackenzie and Neumann 2016). The identification of *C9ORF72* as a common disease gene in ALS and FTD has led them to be currently considered as the ends of a spectrum of related disorders (Renton *et al.* 2014). Although patients may present exclusively with MND or frontotemporal lobar degeneration, in up to 50% of cases there is comorbidity of both disorders (Robberecht and Philips 2013).

The histopathological hallmark features of ALS are the selective loss of motor neurons and the presence of cytoplasmic aggregates of misfolded proteins. Such aggregates are found in most neurodegenerative disorders, including Parkinson's and Alzheimer's disease, where abnormal TDP-43 staining is also frequent (Amador-Ortiz *et al.* 2007; Arai *et al.* 2009). Whether protein aggregates induce or protect from toxicity remains an open question. For the α-synuclein and β-amyloid proteins that form inclusions in Parkinson's and Alzheimer's patients brains, respectively, it is most likely that oligomers of few molecules are more toxic than the insoluble protein aggregates (Haass and Selkoe 2007). In contrast, TDP-43 aggregates have been suggested to be key players in disease pathology, sequestering proteins and/or RNA (Vanden Broeck *et al.* 2014). However, TDP-43 aggregation is not required for neurodegeneration in several animal models (e.g., Iguchi *et al.* 2013), questioning its relevance for ALS pathology. Notwithstanding, several lines of evidence suggest that defective removal of misfolded proteins by the ubiquitin-proteasome system, ER stress and autophagy may be implicated in the disease phenotype (Kanekura *et al.* 2009; Menzies *et al.* 2015). ALS-causing mutations are also linked to changes in other cellular pathways like axonal and mitochondrial dysfunction, oxidative stress, metabolic disturbance, and apoptosis (reviewed by Chen *et al.* 2013). Remarkably, a subgroup of ALS-causative genes can be defined by their connection to RNA metabolic pathways (Table 1). In addition to the RNA-binding proteins (RBPs) TDP-43 and FUS (Rutherford *et al.* 2008; Sreedharan *et al.* 2008; Kwiatkowski *et al.* 2009; Vance *et al.* 2009) discussed above as causing both familial and sporadic ALS, a juvenile form of familial ALS (ALS4) is caused by dominant mutations in the *SENTX* gene, which encodes the DNA–RNA helicase Senataxin (Chen *et al.* 2004). Interestingly, Senataxin is highly related to the immunoglobulin mu binding protein 2 (encoded by the *IGHMBP2*), associated to spinal muscular atrophy with respiratory distress type 1 (Grohmann *et al.* 2001).

Other proteins linked to RNA metabolism implicated in ALS include angiogenin, involved in tRNA and rRNA biogenesis (Greenway *et al.* 2004), Matrin 3 (Johnson *et al.* 2014), and the RBPs hnRNPA1 and hnRNPA2B1 (Kim *et al.* 2013) (Table 1). Although mutations in the *MATR3* gene were originally described as causing an autosomal dominant myopathy (Senderek *et al.* 2009), later clinical re-evaluation of these patients supports the reclassification of their condition as slowly progressive ALS (Johnson *et al.* 2014). Matrin 3 is a regulator of RNA processing (Salton *et al.* 2011; Coelho *et al.* 2015) that interacts directly with TDP-43, FUS, hnRNPA1, and hnRNPA2B1 (Ling *et al.* 2010; Hein *et al.* 2015). Mutations in the later two proteins have been linked to both ALS and multisystem protein disorder (Kim *et al.* 2013). These mutations occur within low-complexity prion-like domains that are intrinsically prone to aggregation and are common in other RBPs, including FUS, TDP-43, and Matrin 3, thereby suggesting an explanation for the existence of 'RBP-related' proteinopathies (Kim *et al.* 2013). However, as mentioned above, several lines of evidence suggest that the formation of protein aggregates is not a requirement for MND.

Although the normal function of the protein encoded by the *C9ORF72* gene remains unclear, several lines of evidence point to connections between the disease-causing mutation and RNA metabolism. Three hypotheses are currently under debate to explain how the intronic hexanucleotide expansion repeat in the *C9ORF72* gene reported in 2011 (DeJesus-Hernandez *et al.* 2011; Renton *et al.* 2011) leads to MND. These include loss-of-function of the C9orf72 protein, a toxic RNA gain-of-function and a toxic effect from di-peptide repeat molecules synthesized by aberrant translation of repeat-containing RNAs (recently reviewed by Gitler and

Table 1 RNA metabolism genes and mutations linked to motor neuron diseases

Protein	HUGO gene ID	Main functions/protein domains	Disease	Inheritance/mutation	REFS
SMN	SMN1	RNP assembly chaperone; snRNP assembly; mRNP transport/Tudor domain	SMA; sALS	AR; total and partial deletions and point mutations	Lefebvre *et al.* (1995) and Burghes and Beattie (2009)
TDP-43	TARDBP	RBP; regulator of splicing, translation and mRNA stability/ RRM; Gly-rich LCD	ALS; FTD	AD/AR; point mutations	Rutherford *et al.* (2008) and Sreedharan *et al.* (2008)
FUS	FUS	RBP; regulator of splicing, translation and mRNA stability/ RRM; Gly-rich LCD	ALS; FTD	AR; point mutations	Kwiatkowski *et al.* (2009) and Vance *et al.* (2009)
Senataxin	SENTX	DNA-RNA helicase; regulator of transcription and genomic stability/ helicase	fALS; AOA2; ADSMA	AD (fALS, ADSMA); AR (AO2)	Chen *et al.* (2004); Moreira *et al.* (2004) and Rudnik-Schöneborn *et al.* (2012)
Angiogenin	ANG	RNA endonuclease; tRNA processing/endonuclease	ALS	AD	Greenway *et al.* (2004)
C9ORF72	C9ORF72	Unclear function; GEF	ALS; FTD	AD; IHER	DeJesus-Hernandez *et al.* (2011) and Renton *et al.* (2011)
hnRNPA1	HRPA1	RBP; regulator of splicing, translation, and mRNA stability/ RRM; Gly-rich LCD	MSP; ALS	AD; missense	Kim *et al.* (2013)
hnRNPA2/B1	HRPA2	RBP; regulator of splicing, translation. and mRNA stability/ RRM; Gly-rich LCD	MSP; ALS	AD; missense	Kim *et al.* (2013)
Matrin 3	MATR3	RBP; nuclear matrix protein, regulator of splicing, and mRNA stability/RRM; Gly-rich LCD	ALS	AD; missense	Johnson *et al.* (2014)

AD, autosomal dominant; ADSMA, autosomal dominant spinal muscular atrophy; AR, autosomal recessive; fALS, familial ALS; FTD, frontotemporal dementia; GEF, GDP/GTP exchange factor; Gly-rich LCD, glycine-rich low complexity domain; IHER, intronic hexanucleotide expansion repeat; MSP, multisystem proteinopathy; RBP, RNA-binding protein; mRNP, messenger ribonucleoprotein particle; snRNP, small nuclear ribonucleoprotein particle; RRM, RNA recognition motif; sALS, sporadic ALS.

Tsuiji 2016). The RNA-dependent toxicity has been proposed to occur through sequestration of RBPs, including TDP-43, potentially disrupting RNA processing (Cooper-Knock *et al.* 2012; Lee *et al.* 2013). The di-peptide repeat proteins formed by repeat-associated non-AUG translation of mutant transcripts have also been shown to interfere with pre-mRNA splicing and ribosomal RNA biogenesis (Kwon *et al.* 2014). Interestingly, both repeat RNAs and di-peptide repeat proteins have independently been implicated in the disruption of nucleo-cytoplasmic trafficking (Freibaum *et al.* 2015; Jovičić *et al.* 2015; Zhang *et al.* 2015).

As mentioned above, *C9ORF72* mutations are the prevailing cause of both ALS and FTD. Several studies report a possible association with other neurodegenerative disorders (Liu *et al.* 2014). However, a recent survey of SMA patients suggests that C9ORF72 variants do not constitute a causal or predisposing factor for this MND (Alías *et al.* 2014).

Spinal muscular atrophies are a heterogeneous group of neurodegenerative disorders characterized by the degeneration of anterior horn cells of the spinal cord, leading to progressive symmetric muscle weakness and atrophy (Simic 2008). We refer to the most common form of this group of disorders as SMA (reviewed by Monani 2005). SMA is caused by hereditary recessive mutations in the survival of motor neuron 1 (*SMN1*) gene, located in the small arm of chromosome 5 (Lefebvre *et al.* 1995). Although rare compared to ALS, SMA is the second most common autosomal recessive disorder after cystic fibrosis, with a pan-ethnic carrier frequency of 1 : 54 (Sugarman *et al.* 2012). It is also the most common cause of hereditary childhood death, usually leading to death within the first year of life (Monani 2005). However, juvenile and adult onset forms also exist. In fact, there are four distinct forms of the disease described: SMA1 (OMIM #253300), SMA2 (OMIM #253550), SMA3 (OMIM #253400), and SMA4 (OMIM #271150), clinically defined by an increase in the age of onset and a decrease in the severity of symptoms. These various forms of SMA are the consequence of the presence of varying copy numbers of a strong disease-modifying gene in the human genome: *SMN2* (Taylor *et al.* 1998). *SMN2* appeared as the

consequence of an inverted duplication of a 500 kb region of the human chromosome 5 containing *SMN1* and some of its neighboring genes (Lorson *et al.* 2010). Compared to *SMN1*, this gene contains a single silent substitution in exon 7 that reduces the splicing efficiency, resulting in a reduced ability to support the synthesis of full length SMN protein (Lorson *et al.* 1999; Monani *et al.* 1999). While in healthy individuals *SMN2* is non-essential, the presence of two copies of this gene is required for viable birth in individuals with homozygous mutations in *SMN1*, and increased copy numbers correlate with decreased disease severity (Taylor *et al.* 1998).

SMA is the paradigm for 'RNA metabolism' associated diseases. As discussed in more detail in the next sections, the SMN protein, encoded by the disease-causing gene, is a critical component of the pathway required for the biogenesis of the splicing machinery. In all animal and yeast models tested so far, SMN is required for cell survival. In humans, the presence of the *SMN2* gene supports just enough SMN protein synthesis to allow for embryonic and fetal development, with conspicuous symptoms usually appearing within a few weeks after birth in SMA1 patients. While the primary pathology is neurodegeneration, it has become increasingly clear that other tissues are also affected, especially in severely affected patients (Shababi *et al.* 2013). Although the reasons for the strong motor neuron phenotype in SMA remain elusive, the prevalence of 'RNA metabolism' among the primary functions of ALS-causing genes suggests a common theme in MND. Among rarer motor neuron diseases this connection is also observed, with reported mutations in *IGHMBP2*, *EXOSC3* (part of the RNA exosome complex), *GLE1* (involved in nuclear RNA export) and, strikingly, *SETX* (reviewed by Peeters *et al.* 2014).

Recently, several publications have revealed strong molecular links between the four key SMA and ALS proteins involved in RNA metabolism, SMN, TDP-43, FUS, and Senataxin. This provides new support for the hypothesis of a shared pathomechanism linked to transcriptome homeostasis in MND, setting the stage for the present review.

Structure and mutant forms of TDP-43, FUS, SMN and Senataxin

TDP-43, or TAR DNA-binding protein with 43 kDa, was originally identified by its ability to bind single-stranded (ss) DNA from the HIV promotor and inhibit its transcription (Ou *et al.* 1995). TDP-43 has two RNA recognition motifs that determine a binding preference for $(UG)_5$ motifs, in addition to binding to TG sequences in ssDNA. It additionally features a C-terminal glycine-rich region that is typical of the heterogeneous nuclear ribonucleoparticle protein (hnRNP) family and is reported to function as a protein–protein interaction domain (Singh and Valcárcel 2005; Fig. 1a). As mentioned above, hyperphosphorylated TDP-43 was found to be a major component of intracellular inclusions in many cases of sporadic and familial ALS and FTLD (Neumann *et al.* 2006). Numerous ALS-linked mutations in the *TARDP* gene have been reported, most of them dominant missense substitutions affecting the C-terminal domain (Lagier-Tourenne *et al.* 2010). However, it is presently unclear whether these mutations are a true gain-of-function or a loss-of-function through dominant-negative effects (Vanden Broeck *et al.* 2014).

Even though it is often referred to as an hnRNP protein, FUS belongs to the TET protein family, which includes the EWS transcriptional regulator and the TATA-binding protein-associated factor TAF15 (Tan and Manley 2009). These three proteins share a similar structure characterized by an N-terminal QGSY-rich region, a highly conserved RNA recognition motifs, multiple RGG repeats, which are extensively dimethylated at arginine residues, and a C-terminal zinc finger motif. FUS additionally contains the glycine-rich region typical of hnRNPs (Fig. 1b). FUS has been reported to bind a common GUGGU motif in RNA molecules (Lagier-Tourenne *et al.* 2012), although it also seems to display DNA-binding activity, including binding to telomeric DNA through the motif $(TTAGGG)_4$ (Takahama *et al.* 2009). However, a more recent study profiling *in vivo* FUS RNA-binding sites failed to identify any clear binding motifs in 80% of the detected FUS targets (Nakaya *et al.* 2013). This suggests that FUS has a low sequence specificity in RNA binding. At least 30 ALS-causative mutations have been reported in the *FUS* gene. The majority of them are dominant missense mutations, clustered either along the glycine-rich region or in the final C-terminal 17 amino acids, which constitute the protein's nuclear localization signal (NLS) (Lagier-Tourenne *et al.* 2010; Fig. 1b). Mutated FUS forms cytoplasmic aggregates in lower motor neurons of affected patients (Vance *et al.* 2009). A recent study has further shown that the physiological role of FUS requires forming dynamic liquid-like compartments, in particular at sites of DNA damage (Patel *et al.* 2015). Such compartments are more prone to aberrant liquid-to-solid phase transitions into aggregated states, which are accelerated by ALS-causing mutations. Although this work does not address the reason why aggregates seem to associate to disease states, it provides a detailed biophysical model for their formation based on specific protein characteristics.

In contrast to FUS and TDP-43, SMN does not belong to the RBP/hnRNP family, being characterized by the presence of a Tudor domain. The Tudor domain is a conserved sequence motif that mediates protein–protein interactions through the recognition of methylated arginine residues (Chen *et al.* 2011). Specifically, the SMN Tudor domain has been shown to bind to RG-rich motifs commonly found in RBPs, depending on the symmetrical dimethylation of specific arginine residues. In addition, SMN presents two dimerization domains required for its function as an

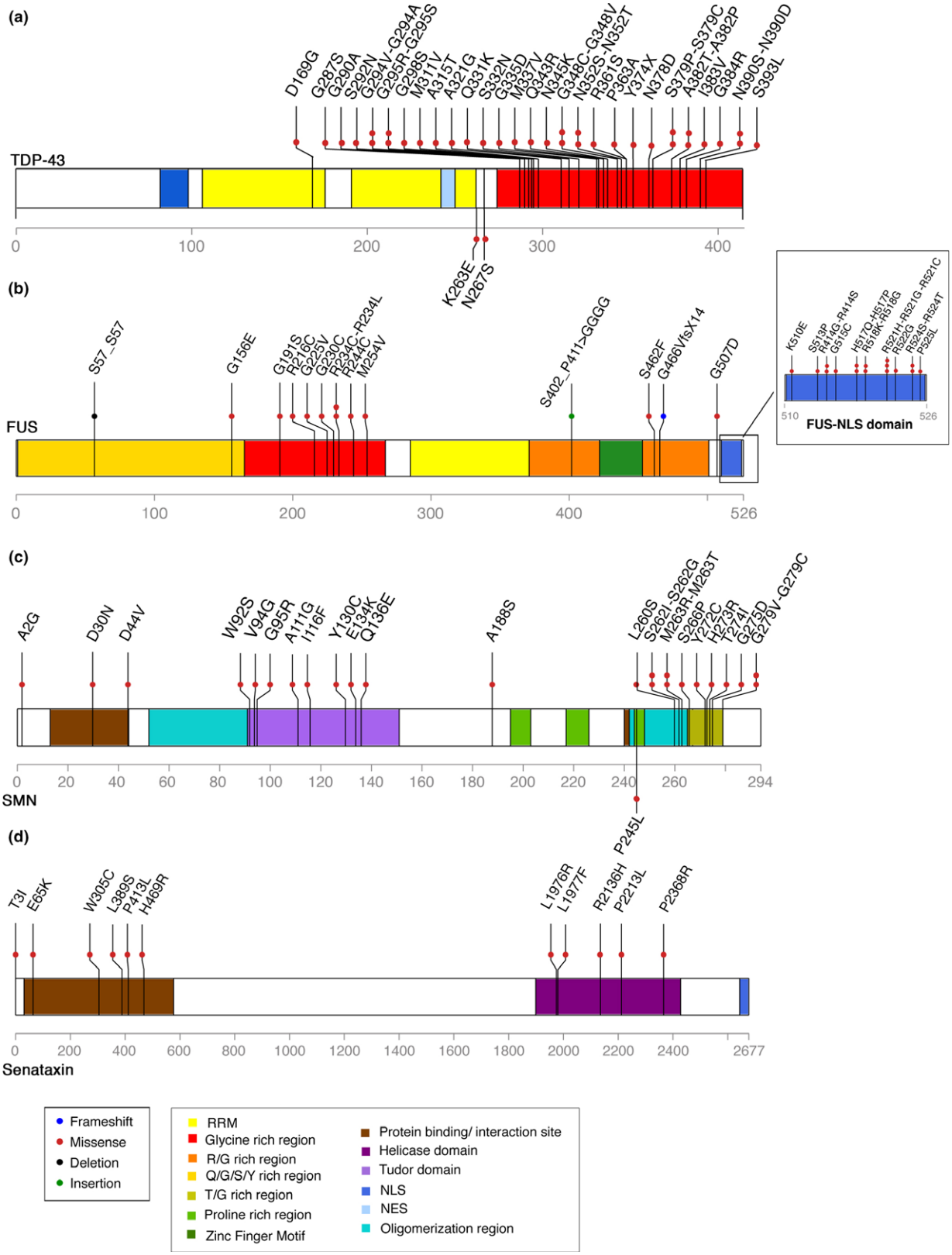

Fig. 1 Structure and function of (a) TDP-43, (b) FUS, (c) SMN, and (d) Senataxin. Figure depicts domain structure and reported disease-causing point mutations in the four proteins. Adapted from (Rossoll and Bassell 2009; Lagier-Tourenne *et al.* 2010; Bennett and La Spada 2015).

oligomeric complex and a YG box domain, commonly found in RBPs. Although the vast majority of SMA-causing mutations involve large deletions of the SMN gene, several missense mutations affecting the Tudor domain and the dimerization and YG box domains have been reported, highlighting the importance of these regions for the disease phenotype (Burghes and Beattie 2009; Fig. 1c).

Finally, Senataxin belongs to the superfamily I of DNA-RNA helicases. It is a large, 302 kDa protein that contains a N-terminal protein–protein interaction domain and a C terminal domain composed of seven helicase motifs, followed by a NLS (Hashemi *et al.* 2006; Fig. 1d). Three dominant missense substitutions, two in the N-terminal and one in the C-terminal helicase domain have been identified in type 4 ALS patients (Chen *et al.* 2004). Strikingly, the ALS-causing missense variant p.L389S was identified in a family suffering from autosomal dominant proximal spinal muscular atrophy, with more affected siblings presenting a second *SETX* missense mutation of unknown function (V891A) in *trans* (Rudnik-Schöneborn *et al.* 2012). Moreover, homozygous missense mutations in the same protein domains linked to ALS are found in patients suffering from AOA2, which presents as a recessively inherited disorder (Moreira *et al.* 2004). Thus, similar mutations in the *SETX* gene can lead to a variety of diseases with neuromuscular symptoms, likely influenced by the genetic background.

SMN, FUS, TDP-43, and STX are shuttling proteins with multiple functions in RNA processing

The functional diversity typically found within the world of 'RNA metabolism' proteins is well illustrated by the four RNA-associated proteins discussed in this review. All these proteins shuttle between the cytoplasm and the nucleus, participating in the regulation of multiple successive steps in the RNA biogenesis pathway.

At steady-state, FUS and TDP-43 accumulate predominantly in the nucleus where they control transcription, splicing and participate in microRNA biogenesis (reviewed by Lagier-Tourenne *et al.* 2010). In neurons, these proteins also localize to RNA granules and regulate their transport to dendritic or axonal extensions. In addition, FUS and TDP-43 regulate the translation of target neuronal mRNAs (Belly *et al.* 2005; Vance *et al.* 2013). As mentioned above, defective subcellular distribution of these proteins is frequently observed in ALS (Neumann *et al.* 2006; Mackenzie *et al.* 2010). For example, disease-associated mutations in FUS were shown to disrupt the normal nuclear localization of mutant proteins and to concomitantly lead to the cytoplasmic accumulation of pathological aggregates. This has raised the debate as to whether the associated MND results from nuclear depletion of FUS or rather a cytotoxic gain of function.

SMN is also a ubiquitous shuttling protein that localizes predominantly in the cytoplasm, with strong accumulation in nuclear structures termed Cajal bodies (CBs). To date, the main reported function of SMN has been as a chaperone for the assembly of U-rich small nuclear (sn) RNA-protein complexes termed UsnRNPs (Bühler *et al.*, 1999). Indeed, the SMN complex, formed by SMN homodimers, Gemins 2–8, and the Unrip protein, is responsible for the assembly of ribonucleoprotein complexes between the Sm and LSm proteins and UsnRNAs (Battle *et al.* 2006). The UsnRNAs are specifically recognized and delivered to SMN-Gemin2 by Gemin5 (Yong *et al.* 2010). This complex's assembly activity is responsible for generating the core cellular machinery involved in diverse RNA pathways, including pre-mRNA splicing (e.g., the major and minor UsnRNPs), histone mRNA 3′-end processing (the U7snRNP), and cytoplasmic mRNA decay (the LSm complex) (He and Parker 2000; Tisdale *et al.* 2013; Li *et al.* 2014). The SMN complex may additionally be required for the assembly of small nucleolar (sno)RNPs involved in rRNA processing (Jones *et al.*, 2001), as well as in the assembly of a U1snRNA-TAF15 transcriptional complex (Jobert *et al.* 2009). Adding to these functions, SMN was shown to recognize methylation marks in histones (Sabra *et al.* 2013) and to interact with the axonal mRNA transport machinery, namely through direct association with actin-binding proteins and neuronal RBPs containing dimethylated RG domains (reviewed by Fallini *et al.* 2012a).

Senataxin was originally named after its well-characterized yeast orthologue Sen1p, a large DEAxQ-box helicase involved in DNA repair and transcription termination. Early studies on the role of Senataxin in AOA2 cells suggested that it was involved in the DNA damage response, which was proposed to be the critical mechanism underlying neurodegeneration (Suraweera *et al.* 2007). Later studies from the same group revealed that, just like yeast Sen1p, the vertebrate protein was involved in the regulation of transcription termination and splicing (Suraweera *et al.* 2009). Like SMN, FUS, and TDP-43, Senataxin has also been reported to be present in the cytoplasm, thereby suggesting that it may have additional cellular functions, in particular in neurons (Hashemi *et al.* 2006; Vantaggiato *et al.* 2011).

In summary, these four proteins form a heterogeneous group of multifunctional regulators of RNA metabolism processes. The fact that FUS, TDP-43, SMN, and Senataxin are all ubiquitously expressed and participate in core cellular functions is one of the major puzzles regarding the tissue specific nature of their associated diseases. Among these functions, pre-mRNA splicing emerges as a common signature that has captured the interest of the research community, given its relevance for the expression of neuronal genes and critical role in establishing tissue specific gene expression patterns.

Defective splicing: a common mechanism underlying FUS, TDP-43, and SMN-dependent pathologies?

The discovery of ALS-causing mutations in FUS and TDP-43 lead to the hypothesis that these proteins might participate in the splicing of common targets critical for motorneuron function. Several labs have since performed systematic searches for these connections, with somewhat disappointing results (Lagier-Tourenne *et al.* 2012; Rogelj *et al.* 2012; Honda *et al.* 2014). Taking advantage of the recent developments in RNA-seq methods, both the mapping of RNA-binding sites and the assessment of changes in splicing patterns in cells and tissues expressing mutant versions of FUS and TDP-43 have been explored (Table 2). In addition, global changes in gene expression have been characterized (see recent review by Lagier-Tourenne *et al.* 2010). Results show that the RNA-binding profiles of FUS and TDP-43 and corresponding regulated splicing events are quite distinct (Polymenidou *et al.* 2011; Rogelj *et al.* 2012). Nevertheless, global classification of the RNA targets of these proteins reveals an enrichment for genes relevant for neuronal function. This may just reflect the fact that neuronal genes tend to present more complex splicing patterns (Zheng and Black, 2013). RNA-seq of mouse brain and spinal cord tissue after *in vivo* knock-down of FUS and TDP-43 revealed only a small subset of commonly affected genes (Lagier-Tourenne *et al.* 2012). These genes were found to contain longer introns than the mouse genome average, suggesting that these proteins may be required for the processing of long pre-mRNAs. Increased interaction of FUS with long introns was reported independently and suggested to be linked to the need to suppress recognition of cryptic splicing sites (Rogelj *et al.* 2012). More recent studies confirmed a role for TDP-43 in the splicing of long introns (Ling *et al.* 2015).

Table 2 Genomic datasets for amyotrophic lateral sclerosis (ALS)/spinal muscular atrophy (SMA) in the Array Express database

Datasets by disease model

Organism	Assay	Total datasets	Total samples	'ALS'	'SMA'	'TDP-43'	'FUS'	'SEXT'
Human	Microarrays	42	2708	28	6	6	1	1
	RNA-seq	7	187	3	4	0	0	0
	CLIP-seq	13	233	6	1	2	4	1
	ChIP	1	2	0	0	0	1	0
Mouse	Microarrays	45	735	22	10	8	3	1
	RNA-seq	12	272	6	3	2	1	0
	RIP-Chip/CLIP-seq	8	79	4	1	2	1	0
	ChIP	1	9	0	0	0	1	0
Rat	Microarrays	4	42	3	0	1	0	0
	RNA-seq	0	0	0	0	0	0	0
	RIP-seq	1	2	0	0	1	0	0
Drosophila	Microarrays	1	36	1	0	0	0	0
	RNA-seq	3	28	1	1	1	0	0
C. elegans	Microarrays	1	6	0	0	1	0	0
	RNA-seq	0	0	0	0	0	0	0
Total		139	4339	64	24	19	5	2

Datasets by tissue/cell type

Organism	Assay	Spinal cord	Neuronal tissue	Muscle	Blood/lymph	Cultured cells	iPSC	ES cells
Human	Microarrays	6	3	6	2	14	7	0
	RNA-seq	1	0	0	1	1	4	0
	CLIP-seq	0	7	0	0	15	0	0
	ChIP	0	0	0	0	1	0	0
Mouse	Microarrays	19	15	6	0	5	0	0
	RNA-seq	6	2	0	0	3	0	3
	RIP-Chip/CLIP-seq	0	6	0	0	3	0	2
	ChIP	0	1	0	0	0	0	0
Rat	Microarrays	2	2	0	0	0	0	0
	RNA-seq	0	0	0	0	0	0	0
	RIP-seq	0	1	0	0	0	0	0
Total		34	22	12	3	23	11	5

Another study analyzing gene expression and splicing changes in mouse cortical neurons after shRNA depletion of TDP-43 and FUS identified two small sets of ~ 50 transcripts commonly affected by both proteins at the gene expression level or at the splicing level (Honda *et al.* 2014). Given the total number of genes affected by each protein individually, this overlap was only significant for genes showing global changes in expression, which may just reflect a common downstream response by neuronal cells. It remains totally unclear how TDP-43 and FUS knock-down lead to these changes. Of note, the degree of overlap between this study and the previous work by Lagier-Tourenne *et al.* (2012) was minimal.

The exploration of splicing defects in SMA has been a rather obvious pursuit since the discovery of the role of SMN in the biogenesis of small nuclear ribonucleoprotein particle (snRNP) complexes. Once again, the results obtained by different groups have revealed conflicting stories and failed so far to establish clear mechanistic connections between SMN levels, splicing aberrations and motor neuron degeneration. Several reports have provided convincing evidence that a decrease in SMN levels results in cell type specific changes in the snRNP repertoire (reviewed by Chari *et al.* 2009). However, although the relative abundance of different snRNPs changes, there is no reason to assume that the composition of spliceosomes assembled on pre-mRNAs is affected, in particular considering that these complexes remain very abundant in cells. Zhang and co-workers were the first to address the genome-wide impact of low SMN levels in splicing using exon arrays. They reported changes in the splicing patterns of roughly 10% of the expressed genes (Zhang *et al.* 2008). Although this is far from a widespread defect as claimed by the authors, the nature of the observed changes and the fact that they occur preferentially in genes with large numbers of introns speaks in favor of a disruption of general splicing mechanisms. Thus, it was proposed that perturbations in the snRNP repertoire influence the efficiency, rate and fidelity of splicing (Zhang *et al.* 2008). This view was supported by a study demonstrating an increased error rate in splice site pairing in cells from SMA patients (Fox-Walsh and Hertel 2009). Focusing on the fact the snRNA components of the minor spliceosome seem to be the most consistently affected by low SMN levels, Lotti and co-workers presented evidence that genes containing minor spliceosomal introns were specifically affected in different SMA models, proposing that this would disrupt a critical cellular pathway for neuronal development (Lotti *et al.* 2012).

The relevance of splicing aberrations because of an altered snRNP repertoire has been called into question by two different types of data. First, two studies have described SMN mutations that uncouple snRNP biogenesis from the motor neuron phenotype (Carrel *et al.* 2006; Praveen *et al.* 2012). In zebrafish, a truncated SMN protein expected to be non-functional in snRNP assembly was able to rescue motor axon defects, unlike the snRNP-assembly competent hSMN (A111G) mutant (Carrel *et al.* 2006). However, a later study from the same authors showed that higher doses of hSMN (A111G) could in fact rescue the phenotype in both zebrafish and mice, in the presence of low levels of wt SMN (Workman *et al.* 2009). The impact of the truncation mutant in snRNP assembly and SMA mice was not evaluated because of technical difficulties. The observation that disruption of other central players of the snRNP assembly pathway can lead to motor phenotypes in mouse (Winkler *et al.* 2005) and fly (Borg *et al.* 2016) lends support to a role for this pathway in SMA, in line with experiments showing that the motor axon phenotype in SMN-deficient zebrafish embryos can be prevented by co-injection of purified UsnRNPs (Winkler *et al.* 2005). However, the observation in *Drosophila* that the SMN^{T205I} disease mutation generates a phenotype compatible with mild SMA without disrupting snRNP biogenesis or snRNA levels suggests that non-snRNP assembly functions of SMN may also play a critical role in the etiology of the disease, at least in this model organism (Praveen *et al.* 2012). This is supported by a later study from the same lab reporting that three hypomorphic SMN missense mutants do not share the splicing profile changes common to SMN null and snRNP biogenesis mutants (Garcia *et al.* 2016).

Second, the so-called widespread changes in splicing have been proposed to represent late features of the disease (Bäumer *et al.* 2009), or an indirect consequence of delayed neuronal development (Garcia *et al.* 2013). In fact, the analysis of splicing defects in early, asymptomatic stages in a mouse model of SMA revealed a very small number of cell-type-specific splicing changes that did not involve minor spliceosomal introns (Zhang *et al.* 2013). These included altered splicing in a few genes involved in synaptogenesis, which the authors argue could be the basis for the disease phenotype. A striking parallel observation was the up-regulation of the mRNAs encoding the three subunits of the C1q protein complex. C1q is known to be involved in neurodegeneration and synaptic pruning during CNS development (Stevens *et al.* 2007). Interestingly, this complex has also been reported to be up-regulated in mouse models of ALS (Saxena *et al.* 2009). Although C1q up-regulation provides an interesting link between SMN levels and neurodegeneration, its change in expression is difficult to explain based on the splicing disruption hypothesis. Thus, many questions still remain unanswered regarding the primary changes in the motor neuron transcriptome that are triggered by mutations in SMA and ALS causal genes, and how they are connected to the disease phenotype. There is, however, an abundance of evidence supporting the existence of functional interactions between the molecular pathways underlying both disorders.

New connections between ALS and SMA

As discussed above, ALS and SMA are both characterized by a progressive degeneration of motor neurons in the anterior horns of the spinal cord. The classical presentation of ALS is that of a sporadic disorder with onset after the forth decade of life, but with a very heterogeneous clinical presentation, from site and age of onset, familial occurrence, type of motor neuron involvement, extent of extra-motor involvement and disease duration, among others (Swinnen and Robberecht 2014).

In spite of a more consistent clinical presentation as a very severe, hereditary, early childhood disorder, SMA can overlap with the ALS spectrum of manifestations. Indeed, cases of late to adult onset of the disease are known, usually linked to the presence of extra copies of the *SMN2* gene, as discussed above. Interestingly, several authors have reported SMN1 and SMN2 copy numbers as susceptibility factors for sporadic ALS in different populations (Veldink *et al.* 2005; Corcia *et al.* 2012; reviewed by Butchbach 2016). Although there are some conflicting data, a recent study of a large cohort of ALS patients, coupled with a meta-analysis of previous data, reported SMN1 duplications as one of the highest known risk factors for sporadic ALS (Blauw *et al.* 2012). The emerging genetic connections between the two diseases are further underscored by studies in cellular and animal models of ALS that report functional interactions with SMN. Indeed, over-expression of SMN was shown to rescue the axonal growth and branching-defects observed in FUS mutant neurons (Groen *et al.* 2013). Increased SMN levels also improved neuromuscular function and motor neuron survival in mice with expression of a ALS-causing mutation in the SOD1 gene (Turner *et al.* 2014). Finally, a very recent study shows that neuronal over-expression of human SMN delays symptom onset and prolongs survival in mice bearing the TDP-43^{A315T} mutation, by counter-acting both motor neuron degeneration and inflammatory glial cell activation (Perera *et al.* 2016).

Taken together, these studies reveal the existence of shared mechanisms underlying ALS and SMA, which may actually extend to settings that appear unrelated to RNA metabolism, as is the case of SOD1 mutations. Such connections are beginning to be unraveled by both genetic and functional studies and, importantly, through the analysis of the protein interaction profiles of the disease-causing genes, as discussed next.

Protein networks revealing links between SMN, TDP-43, FUS and Senataxin

Mapping of protein–protein interactions provides an efficient way to obtain insights into the cellular functions of a protein. Several studies have directly addressed the protein interaction profiles of SMN, TDP-43, FUS, and Senataxin, revealing direct connections between all of them (Suraweera *et al.* 2009; Ling *et al.* 2010; Yamazaki *et al.* 2012; Sun *et al.* 2015). Data from high-throughput proteomic approaches to systematically map protein–protein interactions in different model organisms are publicly available and can be used to generate a relatively detailed view of the interactomes of these four proteins. We performed a systematic retrieval of the interactions for SMN, TDP-43, FUS, and Senataxin from different data resources and recovered a set of 136 proteins that interact simultaneously with at least two of them (Fig. 2). This analysis revealed a highly interconnected network of interactions between the four proteins, which preserves around 20% of each individual interactome. Noteworthy, the analysis of functional gene ontology terms associated to this network showed nine highly enriched functional clusters linked to: (i) spliceosome assembly; (ii) translational elongation/termination; (iii) RNA processing; (iv) ubiquitin-dependent protein degradation; (v) DNA repair; (vi) rRNA processing; (vii) transcription termination; (viii) microtubule-based movement, and (ix) cell death and apoptosis (Fig. 2). These clusters highlight some expected connections, considering the current knowledge about SMA and ALS (such as spliceosome assembly, rRNA processing, or translation control). More interestingly, they foreshadow the most recent discoveries linking TDP-43, FUS, SMN, and Senataxin to specific aspects of gene expression regulation, including transcription termination and DNA repair, as discussed in the next sections.

A common cellular location for the four proteins: buddies in nuclear bodies

The finding that the SMN protein is involved in the assembly of UsnRNPs led to the early hypothesis that this process is affected in SMA (Liu *et al.* 1997). Indeed, later biochemical studies showed that snRNP assembly is less efficient in cell lines with SMN down-regulation as well as in SMA patient cells (reviewed by Chari *et al.* 2009).

snRNP biogenesis is a complex process that begins with the RNA Polymerase II (Pol II)-dependent transcription of two types of RNAs: mRNAs coding for Sm core and snRNP specific proteins; and non-coding UsnRNAs (with the exception of the U6snRNA, transcribed by RNA Polymerase III). Both types of RNAs are exported to the cytoplasm, where the SMN complex helps to assemble snRNAs and Sm proteins into early snRNPs (Fig. 3). snRNAs, however, accumulate first in a discrete nuclear structure, the Cajal body (CB), where they undergo several processing steps before moving to the cytoplasm (reviewed by Matera and Wang 2014). After assembly in the cytoplasm, snRNPs are re-imported into the nucleus and targeted to CBs, where they are subject to further maturation steps before becoming active in splicing.

In spite of the absence of membranous compartments, the nucleus has a high degree of spatial organization. The nucleoli represent the most prominent and best-known

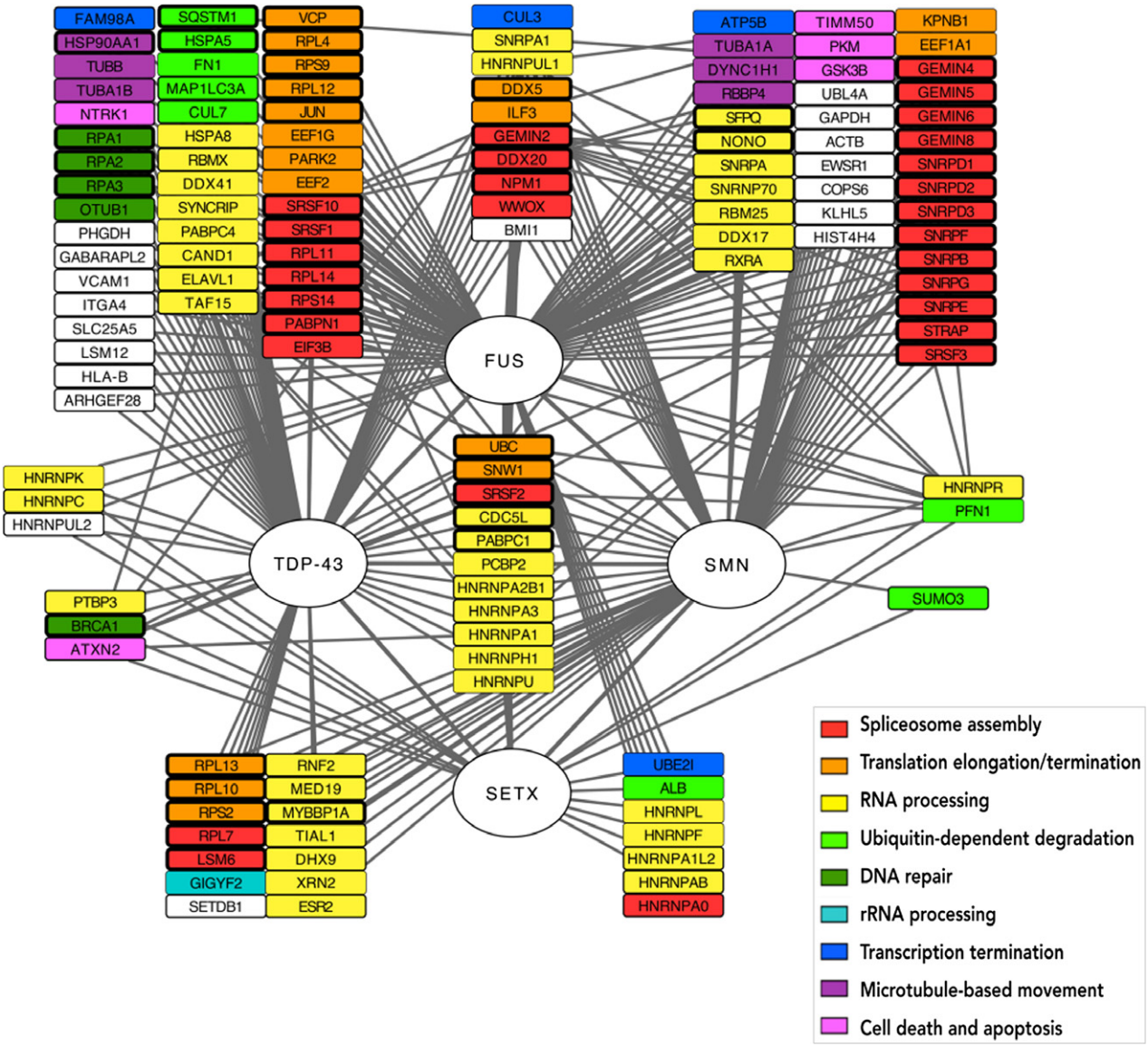

Fig. 2 The FUS, TDP-43, SMN, and Senataxin interactome. Physical interactions for each protein were recovered from GeneMania, IntAct, Mentha and by manual curation of the literature (available as a supplementary data file). Interaction partners shared by at least two of these proteins are presented in the figure, corresponding to a total of 136 proteins. The network was generated using Cytoscape. Functional classification of the interaction network was performed using TopGo for functional enrichment and GeneTerm Linker for clustering of enriched GO terms (Aibar *et al.* 2015). Proteins are color coded according to the corresponding functional cluster, as presented in the legend. Proteins that are part of more than one functional cluster have a thick node border and the color shown represents the functional clusters with the lower enrichment *p*-value.

example of a 'subnuclear' compartment. During the last decades, the existence of a plethora of such structures, often termed 'nuclear bodies', has become apparent. Among them are the CBs, gemini of Cajal bodies (or gems) and histone locus bodies (HLBs) (reviewed by Machyna *et al.* 2013). CBs were first identified by Ramón y Cajal in CNS neurons. Decades later, CBs were also found in non-neuronal cells and referred to as coiled bodies (reviewed by Gall 2000). During the last decades, CB function has been characterized in greater detail. It has been suggested that CBs represent the core of the three above mentioned nuclear bodies, which clearly seem to be functionally related. Indeed, CBs, gems and HLBs are surprisingly dynamic and depending on the cell type and metabolic state, they may or may not be visible or even seem to coalesce with each other into a single body (Machyna *et al.* 2013). CBs are defined by the presence of the protein coilin and are primarily believed to represent way stations in the snRNP biogenesis pathway (Fig. 3; reviewed

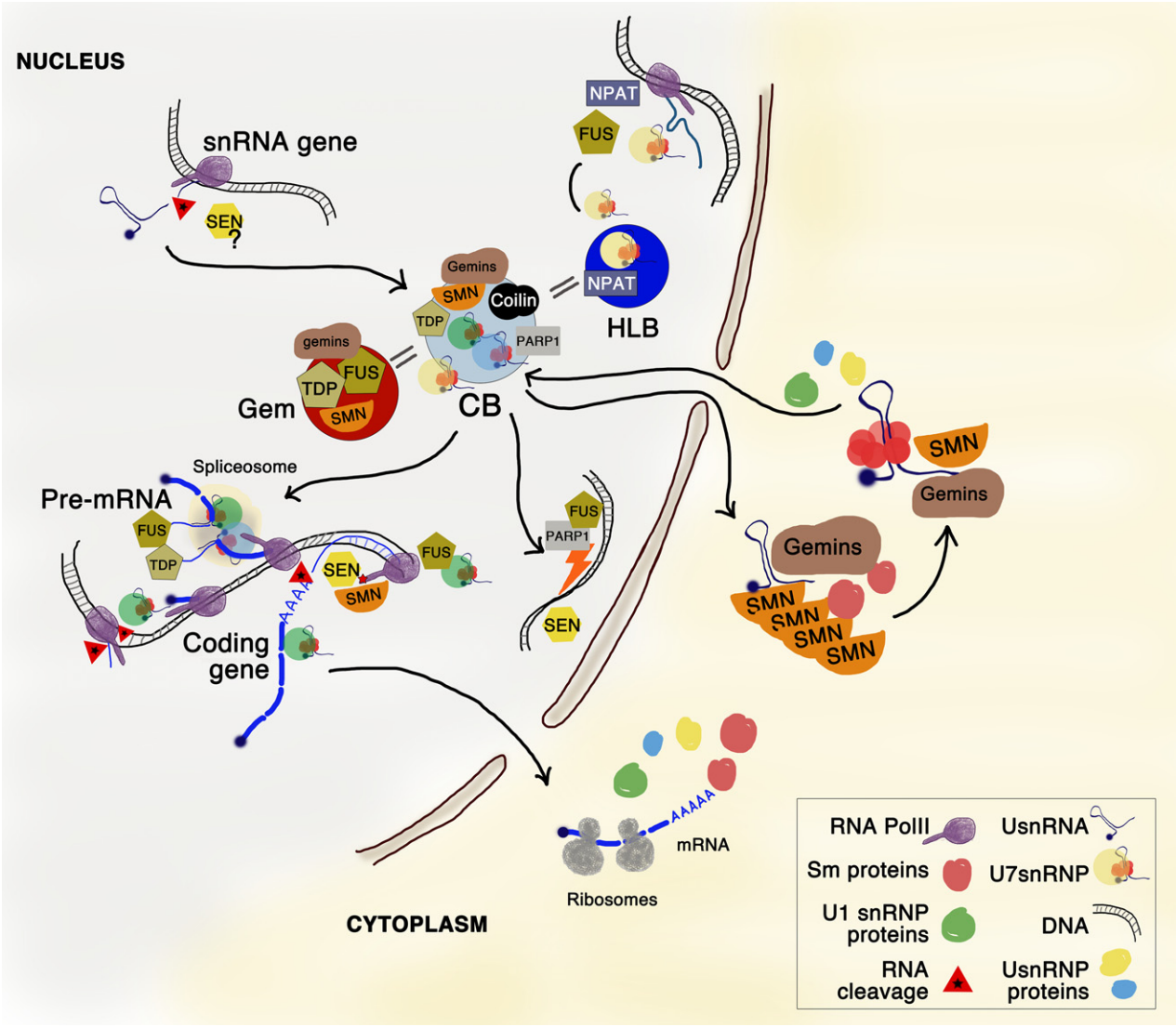

Fig. 3 Connections between TDP-43, FUS, SMN, and Senataxin to gene expression, the snRNP biogenesis, nuclear bodies, and chromatin maintenance and repair. See text for details. CB, Cajal body; HLB, histone locus body.

by Matera and Wang 2014). In addition, they contain several other proteins and RNAs involved in snRNP maturation and recycling after splicing, including the SMN complex. CBs have also been shown to associate with snRNA and histone gene clusters (Frey and Matera 1995), and may contain histone processing and transcription factors, including the U7snRNA and NPAT. However, these proteins often appear to segregate to the related HLBs, which lack SMN and spliceosomal UsnRNPs (Nizami *et al.* 2010). Gems were originally described by the Dreyfuss lab as a distinct nuclear structure from CBs, containing only SMN complex components but not snRNPs (Liu and Dreyfuss 1996). The function of gems remains presently unknown and they seem to coalesce with CBs, in particular in contexts of increased snRNP maturation (Liu and Dreyfuss 1996). Notably, in motor neurons, gems, and CBs predominantly coincide in a single structure (Cioce and Lamond 2005).

In agreement with a role for CBs and gems in snRNP biogenesis, the hallmark cellular phenotype of SMA is a disease severity-dependent reduction in their numbers (Feng *et al.* 2005). Recently, several groups have reported that gems are also disrupted in ALS models, therefore establishing another striking connection between both diseases (reviewed by Cauchi 2014). Both TDP-43 and FUS were found to bind to SMN and accumulate in gems, and to be required for the maintenance of these structures (Yamazaki *et al.* 2012; Ishihara *et al.* 2013; Tsuiji *et al.* 2013). TDP-43 has also been reported to be present in CBs, which likewise seem to be reduced in number in ALS spinal cord MNs (Ishihara *et al.* 2013). The mouse homolog of FUS (Pigpen) was one of the

first CB components to be identified, suggesting that FUS also has connections to this subcellular compartment (Cioce and Lamond 2005). Moreover, the targeting of SMN to CBs and the assembly of gems were found to be disrupted in SOD1 mutant mice, thereby placing these structures at the core of the ALS phenotype (Kariya *et al.* 2012). Additional recent results suggest that the connections between ALS, SMA and the processes that occur in nuclear bodies may actually extend further. One peculiar point of contact is linked to the 3′-end processing of histone mRNAs.

New roles for the proteins together: a tale of histones and DNA

In spite of being transcribed by RNA Pol II, replication-dependent histone mRNAs lack a poly-A tail and present a unique 3′-end structure that requires specific processing and transcription termination events (Marzluff *et al.* 2008). Histone mRNA 3′end processing requires the U7snRNP and seems to occur in association with nuclear bodies, either CBs or HLBs, which assemble near histone gene clusters (Frey and Matera 1995). As could be predicted from SMN's role as a chaperone for snRNP assembly, Tisdale and colleagues recently reported that low levels of SMN affect U7snRNP biogenesis and, in consequence, histone mRNA processing (Tisdale *et al.* 2013). This observation led to the suggestion that defects in histone biogenesis could contribute to the etiology of SMA. This hypothesis could provide an explanation for the increased susceptibility of SMA patient fibroblasts to genotoxic stress (Tisdale *et al.* 2013). In a recent twist to this story, Raczynska *et al.* reported that FUS is also involved in histone mRNA biogenesis. FUS interacts with the U7snRNP and the histone transcriptional co-activator NPAT, leading both to increased transcription and processing of histone mRNAs (Raczynska *et al.* 2015). Thus, FUS and SMN seem to be connected to the processing of these mRNAs in a context that likely involves CBs and their associated structures (Fig. 3).

CBs have been linked to several additional cellular functions, including telomere maintenance and the cellular response to stress (Machyna *et al.* 2013), two functions that have also been reported for FUS (reviewed by Sama *et al.* 2014). The recruitment of FUS to sites of DNA damage was recently shown to be dependent on the CB component PARP1 (Rulten *et al.* 2014). PARP1, a well known regulator of the DNA damage response (Schreiber *et al.* 2006), is required for the assembly and maintenance of CBs and was proposed to act by assisting the shuttling of proteins from CBs to chromatin and back (Kotova *et al.* 2009).

It is noteworthy that CBs respond to many cellular stimuli, including DNA damage, by changing their number and structure. A recent and intriguing publication suggests that CBs may orchestrate the genome-wide clustering of snRNA, histone, spliceosome, and DNA damage repair-related genes (Wang *et al.* 2016). CBs were thus proposed to function as genome organizing centers with a widespread influence on DNA and RNA metabolism (Wang *et al.* 2016). The presence of many ALS- and SMA-related proteins within CBs and in their associated compartments gems and HLBs thus suggests an intriguing connection between MND, genome homeostasis, and RNA metabolism.

Common association with small nuclear ribonucleoproteins

The reported disruption of gems observed in cells with mutant TDP-43 and FUS suggested that snRNP biogenesis could also be affected in ALS. In fact, reduction in TDP-43 was shown to cause a down-regulation of SMN and gemins, together with alterations in snRNP abundance (Ishihara *et al.* 2013; Tsuiji *et al.* 2013). The changes in snRNP abundance are, however, distinct from those found in SMA cells. In fact, although reported results vary with cell type, the abundance of several UsnRNAs seems predominantly to increase in cells with low levels or mutant TDP-43 (Ishihara *et al.* 2013; Tsuiji *et al.* 2013). Notably, this increase in snRNA levels is coupled to their subcellular mislocalization, which could be consistent with a functional reduction in snRNP availability (Tsuiji *et al.* 2013). Thus, the functional impact of abnormal TDP-43 function on snRNP dependent activities may turn out to be similar to the one observed in SMA cells.

Similar to TDP-43, FUS was also shown to interfere with the cellular availability of snRNPs (Gerbino *et al.* 2013; Sun *et al.* 2015). Analysis of fibroblasts derived from ALS patients revealed a mutation-dependent impact of FUS on snRNA levels (Sun *et al.* 2015). Thus, the H517Q substitution and the frameshift mutation M511Nfs*6 resulted in a reduction in snRNA levels. In contrast, the R521G substitution, also localized in the NLS domain, resulted in an increased abundance of several major class snRNAs, accompanied by the down-regulation of minor class snRNAs U4atac and U6atac. Mouse spinal cord MNs revealed yet another distinct profile, with significant reduction in the levels of the U1 and U11 snRNAs (Sun *et al.* 2015). These results are reminiscent of the cell-type-dependent effects observed for TDP-43 mutations, while confirming that both proteins interfere with snRNP biogenesis. This fact brings the impact of ALS-causing mutations in FUS and TDP-43 closer to the SMA phenotype. It remains presently unclear whether common signatures indicative of a general perturbation of the snRNP machinery can be found in both diseases.

Analysis of the FUS interactome identified the U1 snRNP as one of its main binding partners, in addition to SMN (Yamazaki *et al.* 2012; Sun *et al.* 2015). Components of the U1snRNP Sm core particle were found to be mislocalized in cells expressing the ALS-causing R495X mutation in the FUS NLS. Moreover, and U1snRNP knockdowns in different models resulted in similar phenotypes observed to be

induced by FUS mutations, exposing a strong functional connection between the two (Gerbino *et al.* 2013; Yu and Reed 2015). Notably, U1 snRNPs are known to be involved in additional, non-spliceosomal functions (West 2012). In agreement with this, their cellular abundance is estimated to be 2–4 fold higher than that of other UsnRNPs (Baserga and Steitz 1993). This imbalance was recently shown to be achieved thanks to an alternative assembly pathway in which the U1snRNA is directly presented to the SMN-Gemin2 complex by the U1-70K protein rather than Gemin 5 (So *et al.* 2016). This recent study opens the possibility that multiple RBPs may act to present different RNA molecules to the SMN-Gemin2 complex, which could thus function more broadly as an RNP-exchanger, facilitating the assembly of RNPs from multiple donors and acceptors.

The first evidence for additional cellular functions for the U1snRNP came from the elucidation of the mechanism by which the presence of an intron near the transcription start site (TSS) leads to an increase in transcription initiation (Furger *et al.* 2002). The presence of a promotor-proximal 5′ splice site allows the U1 snRNP and TFIIH transcription factors to interact, enhancing the formation of the first nucleotide bond after initiation (Kwek *et al.* 2002). Alternative U1 snRNP complexes involving the transcriptional regulator TAF15 have also been reported (Jobert *et al.* 2009). In agreement with a splicing-independent function in gene expression, U1snRNP is the only UsnRNP found to associate with the loci of intronless genes (Spiluttini *et al.* 2010).

The functional versatility of U1snRNP extends well beyond transcriptional initiation. Binding of this complex to pre-mRNAs has been known for a long time to inhibit the use of poly-A signals (Gunderson *et al.* 1998). The relevance of this inhibitory interaction is actually manifold. First, it was shown to be essential to support transcription across long introns (Alexander *et al.* 2010; Kaida *et al.* 2010). poly-A signals are very frequent in intronic sequences and binding of U1snRNP prevents premature cleavage of the pre-mRNA by suppressing their recognition. Second, reduction in U1snRNP abundance under specific cellular conditions can result in a regulatory switch promoting the selection of upstream poly-A sites, thus leading to a shortening of 3′UTR sequences (Berg *et al.* 2012). Interestingly, this phenomenon seems to be particularly important in neuronal activation. Finally, the interaction between poly-A signals and U1snRNP-binding sites was shown to have a critical role in regulating transcription directionality (Almada *et al.* 2013). Genome-wide analysis of mammalian gene expression has revealed that transcription initiates bi-directionally at CpG promotors. The absence or presence of U1-binding sites on each side of the promotor establishes the distinction between the productive and non-productive direction of transcription. This ensures the quick cleavage and degradation of one of the products, while the other remains protected. Thus, the U1snRNP seems to be a critical element for proper transcriptional balance across the whole genome.

Considering the interactions between SMN, FUS and the U1snRNP described above, this new perspective on the function of U1snRNP points to a potential role for altered transcription initiation and termination in SMA and ALS. In agreement with this, FUS was recently shown to influence mRNA length in a way very similar to the U1snRNP, again with the authors proposing a preponderant role for this regulatory mechanism in neuronal gene expression (Masuda *et al.* 2015). This observation was further expanded by a report from the Reed lab showing that FUS is required to mediate interactions between RNA Pol II and the U1snRNP (Yu and Reed 2015). Thus, U1snRNP splicing-independent functions may contribute to changes in transcription control critical for motor neuron degeneration. Just as an unexpected twist in the plot, the recent identification of a novel critical role for SMN and Senataxin in Pol II transcription termination further extends this perspective and provides an exciting new connection between ALS and SMA.

New components in the transcription molecular machinery

Transcription by Pol II has been known for a long time to be highly interconnected to RNA processing events. This occurs in large part through a complex signaling and docking platform provided by the Pol II C-terminal domain (CTD) (reviewed by Bentley 2014). These connections extend from all steps of transcription (initiation, elongation, and termination) down to events required for the formation of a mature mRNA, including the packaging of the molecule into an mRNP-export competent complex. The association of the nascent RNA molecule to RBPs occurs as soon as it emerges from the Pol II complex, with the CTD ensuring the immediate local availability of the appropriate proteins. This is achieved through the dynamic modulation of CTD-protein interactions by an elaborate flow of post-transcriptional modifications that reflect the ongoing transcription stage (Bentley 2014). The association of RBPs with RNA has additionally been shown to be critically important to avoid the invasion of the template DNA-strand by the nascent transcript. This would lead to the formation of DNA/RNA hybrid structures known as R-loops (reviewed by Santos-Pereira and Aguilera 2015). Uncontrolled R-loop formation occurs mostly over G-rich regions potentially leading to increased DNA damage and genome instability (Santos-Pereira and Aguilera 2015). However, recent results have shown that R-loops have a critical function at the end of many mammalian genes, by inducing Pol II pausing and supporting transcription termination (Skourti-Stathaki *et al.* 2011).

Although the 3′ ends of mRNA molecules are formed by cleavage at poly-A sites, Pol II transcription continues

beyond this point until the enzyme is derailed by the 5′-to-3′ Xrn2 exonuclease (reviewed by Porrua and Libri 2015). Xrn2 latches on to the free, uncapped 5′ end of the cleaved RNA transcript, degrading it and leading to Pol II termination. This process is facilitated by Pol II pausing at G-rich regions downstream of the Poly-A site. Interestingly, Pol II pausing was shown to occur as a result of R-loop formation, in the absence of which transcription will continue, leading to read-through intergenic products (Skourti-Stathaki *et al.* 2011). Proper transcription termination further requires the resolution of these R-loops, allowing Xrn2 to access the RNA transcript. This was shown to be performed by Senataxin (Skourti-Stathaki *et al.* 2011), thus explaining the previous observations that implicated this protein in the termination of transcription (Suraweera *et al.* 2009). The recent publication of Zhao and colleagues sheds further light onto this story by identifying SMN as the intermediary that connects Senataxin to the Pol II CTD (Zhao *et al.* 2016). SMN, already known to functionally interact with the Pol II CTD (Pellizzoni *et al.* 2001), turns out to bind specifically to the di-methylated form of the arginine residue (R1810) in Pol II, recruiting Senataxin to elongating complexes. Knock-down of SMN or Senataxin results in global changes in the association of Pol II to chromatin, which accumulates in the termination region of active genes (Zhao *et al.* 2016). This occurs in parallel with the accumulation of R-loops at the end of genes, affecting the efficiency of termination. Reduced transcription initiation was also reported in these knock-down experiments, which may be linked to the observation that R-loops are also found at the promotor region, immediately after the TSS (Ginno *et al.* 2012). Interestingly, mutation of the Pol II CTD R1810 to an alanine prevents its methylation and was found to specifically result in altered snRNA expression in human cells (Zhao *et al.* 2016).

The relevance of R-loops in the context of ALS had already been proposed and discussed by several authors, with particular emphasis on the connections to genome stability and genotoxic stress (Salvi and Mekhail 2015). Zhao and co-workers recent report reveals that SMA patient-derived cells display similar phenotypes to the ones observed in Senataxin knock-down cells, as would be expected from this new SMN function (Zhao *et al.* 2016). These results provide novel insights regarding the mechanisms underlying motor neuron degeneration in SMA and ALS4, highlighting unsuspected connections between the two diseases. Furthermore, they significantly expand a model of highly inter-twined regulation of RNA synthesis and genome homeostasis involving transient interactions and complexes formed between TDP-43, FUS, Senataxin and SMN. Although not discussed in this review, these connections clearly extend to cytoplasmic functions through the control of axonal RNA-trafficking, translation and the formation of structures like RNA stress-granules (see, for example, Fallini *et al.* 2012a,b; Alami *et al.* 2014; Groen *et al.* 2013). It is likely that such functions also play an important role in the motor neuron-specific effects of disease-causing mutations in these proteins.

Enter omics: genomic and transcriptomic data analysis to reveal new links

The recent insights on the functional interactions between FUS, TDP-43, SMN and Senataxin strongly suggest the existence of common pathway(s) for motor neuron degeneration connected to RNA-transcriptome homeostasis. To date, the reason for the cell type specificity of the ALS/SMA phenotype remains elusive. In particular, whether there are specific genes with altered expression leading to MND or whether motor neurons are responding with high sensitivity to a disruption in transcriptome homeostasis, coupled to alterations in cytoplasmic functions, is presently unclear. We propose that a comparative study of the transcriptome changes present in different cell types with SMN, TDP-43, FUS, and Senataxin disease-causing mutations may assist in the identification of common disease pathways and thus contribute to unravel the specific mechanisms underlying motor neuron death. Several public datasets from human patient tissues and cells, as well as from mouse and rat models of these diseases, are currently available (Table 2). These data provide a starting point for an integrative approach to ALS and SMA, which could at the same time explore the relevance of the links that have recently emerged. Analysis of genomic data is strongly influenced by the computational algorithms, which may be more suited to address certain expected changes in detriment to others. For example, most approaches aimed at quantifying gene or splicing isoform levels will not be efficient at assessing altered transcription termination. In addition, meta-analysis studies of different mutant backgrounds and the integration of results focusing on different approaches – that is, chromatin association versus transcript levels – has not been pursued in many cases. Thus, the reanalysis of these datasets in light of the emerging new connections between ALS and SMA may provide interesting insights into the mechanisms underlying motor neuron degeneration. Considering the existence of several animal models that mimic both the SMA and ALS motor neuron degeneration phenotype, the possibility of comparing cross-species data may provide an additional guide to the identification of common critical pathways. However, studies using matched tissues and conditions for all mutations are lacking and would provide a critical contribution for the implementation of this kind of cross-disease approach.

Conclusions: ALS and SMA diseases disrupt RNA-transcriptome homeostasis

The identification of disease-causing mutations and the downstream molecular mechanisms that link ALS and SMA to the ensuing phenotype can provide the basis for the development of effective therapeutic approaches. However,

moving from the disease-causing gene to mechanism can be quite a challenging task, in particular in contexts were the mutation of an ubiquitously expressed protein leads to cell type specific phenotypes, as is the case with SMA and ALS. Accordingly, more than one decade after the identification of the underlying disease-associated genes, we are still far from understanding how their encoded proteins are connected to motor neuron death.

During the last few years, several lines of evidence have suggested the existence of shared molecular pathways between ALS and SMA. These come from genetic studies reporting the association between aberrant SMN copy numbers and ALS predisposition, to an increasing number of reported interactions between disease-causing genes. In spite of evident commonalities, these diseases have been traditionally studied separately, probably because of the specificity of the clinical manifestations of their most frequent forms. The fact that SMA usually appears as an early childhood hereditary disorder, while ALS presents as a late onset, progressive adult neurodegeneration, most frequently of sporadic nature, has probably been the main driving force in this divide. We propose that the most recent evidence connecting two of the ALS and the SMA disease-causing genes supports the hypothesis of common critical pathways linked to RNA-transcriptome homeostasis underlying motor neuron degeneration in these diseases. Comparative studies, in particular those using high-throughput approaches, may provide the focus required for their identification.

Acknowledgments and conflict of interest disclosure

This work is part of an EU Joint Programme – Neurodegenerative Disease Research (JPND) project with the acronym *'Fly-SMALS'*. The project is supported through the following funding organisations under the aegis of JPND – www.jpnd.eu: France, Agence Nationale de la Recherche; Germany, Bundesministerium für Bildung und Forschung (BMBF); Portugal, Fundação para a Ciência e a Tecnologia and Spain, Instituto de Salud Carlos III (ISCIII). Associated to the JPND, the group of JDLR was funded for this work by the ISCIII and FEDER through projects AC14/00024 and PI15/00328. Work in MGC's group was supported by the grant JPND-CD/0002/2013 and FCT/MCTES/PIDDAC research center grant to BioISI UID/MULTI/04046/2013. Work in F.B.'s group is supported by the ANR (through the RNAGRIMP research grant and the 'Investments for the Future' LABEX SIGNALIFE program # ANR-11-LABX-0028-01). Work on ALS in Joachim Weis's group was supported by funding from BMBF, IZKF Aachen (N7–4), the Deutsche Gesellschaft für Muskelkranke (DGM) and the Initiative Therapieforschung ALS. Jörg B. Schulz is Editor-in-Chief of the Journal of Neurochemistry.

All experiments were conducted in compliance with the ARRIVE guidelines.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Data S1. Physical interactome of FUS, TDP-43, SMN and Senataxin and list of network nodes and edges used to generate Fig. 2.

Data S2. Annotated list of available public transcriptome datasets for ALS and SMA research summarized in Table 2.

References

Aibar S., Fontanillo C., Droste C. and De Las Rivas J. (2015) Functional Gene Networks: R/Bioc package to generate and analyse gene networks derived from functional enrichment and clustering. *Bioinformatics* **31**, 1686–1688.

Alami N. H., Smith R. B., Carrasco M. A. *et al.* (2014) Axonal transport of TDP-43 mRNA granules is impaired by ALS-causing mutations. *Neuron* **81**, 536–543.

Alexander M. R., Wheatley A. K., Center R. J. and Purcell D. F. J. (2010) Efficient transcription through an intron requires the binding of an Sm-type U1 snRNP with intact stem loop II to the splice donor. *Nucleic Acids Res.* **38**, 3041–3053.

Alías L., Bernal S., Barceló M. J., Martínez-Hernández R., Martínez E., Baiget M. and Tizzano E. F. (2014) Analysis of the C9orf72 gene in spinal muscular atrophy patients. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* **15**, 563–568.

Almada A. E., Wu X., Kriz A. J., Burge C. B. and Sharp P. A. (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* **499**, 360–363.

Amador-Ortiz C., Lin W.-L., Ahmed Z., Personett D., Davies P., Duara R., Graff-Radford N. R., Hutton M. L. and Dickson D. W. (2007) TDP-43 immunoreactivity in hippocampal sclerosis and Alzheimer's disease. *Ann. Neurol.* **61**, 435–445.

Arai T., Mackenzie I. R. A., Hasegawa M., Nonoka T., Niizato K., Tsuchiya K., Iritani S., Onaya M. and Akiyama H. (2009) Phosphorylated TDP-43 in Alzheimer's disease and dementia with Lewy bodies. *Acta Neuropathol.* **117**, 125–136.

Baserga S. J. and Steitz J. A. (1993) The diverse world of small ribonucleoproteins. *Cold Spring Harbor Monograph Archive* **24**, 359–381.

Battle D. J., Kasim M., Yong J., Lotti F., Lau C.-K., Mouaikel J., Zhang Z., Han K., Wan L. and Dreyfuss G. (2006) The SMN complex: an assembly machine for RNPs. *Cold Spring Harb. Symp. Quant. Biol.* **71**, 313–320.

Bäumer D., Lee S., Nicholson G., Davies J. L., Parkinson N. J., Murray L. M., Gillingwater T. H., Ansorge O., Davies K. E. and Talbot K. (2009) Alternative splicing events are a late feature of pathology in a mouse model of spinal muscular atrophy. *PLoS Genet.* **5**, e1000773.

Belly A., Moreau-Gachelin F., Sadoul R. and Goldberg Y. (2005) Delocalization of the multifunctional RNA splicing factor TLS/FUS in hippocampal neurones: exclusion from the nucleus and accumulation in dendritic granules and spine heads. *Neurosci. Lett.* **379**, 152–157.

Bennett C. L. and La Spada A. R. (2015) Unwinding the role of senataxin in neurodegeneration. *Discov Med* **19**, 127–136.

Bentley D. L. (2014) Coupling mRNA processing with transcription in time and space. *Nat. Rev. Genet.* **15**, 163–175.

Berg M. G., Singh L. N., Younis I. *et al.* (2012) U1 snRNP determines mRNA length and regulates isoform expression. *Cell* **150**, 53–64.

Blauw H. M., Barnes C. P., van Vught P. W. J., van Rheenen W., Verheul M., Cuppen E., Veldink J. H. and Van den Berg L. H. (2012) SMN1 gene duplications are associated with sporadic ALS. *Neurology* **78**, 776–780.

Borg R. M., Fenech Salerno B., Vassallo N., Bordonné R. and Cauchi R. J. (2016) Disruption of snRNP biogenesis factors Tgs1 and pICln induces phenotypes that mirror aspects of SMN-Gemins complex

perturbation in *Drosophila*, providing new insights into spinal muscular atrophy. *Neurobiol. Dis.* **94**, 245–258.

Bühler D., Raker V., Lührmann R. and Fischer U. (1999) Essential Role for the Tudor Domain of SMN in Spliceosomal U snRNP Assembly: Implications for Spinal Muscular Atrophy. *Human Molecular Genetics* **8**, 2351–2357.

Burghes A. H. M. and Beattie C. E. (2009) Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick? *Nat. Rev. Neurosci.* **10**, 597–609.

Butchbach M. E. R. (2016) Copy number variations in the survival motor neuron genes: implications for spinal muscular atrophy and other neurodegenerative diseases. *Front Mol Biosci* **3**, 7.

Carrel T. L., McWhorter M. L., Workman E., Zhang H., Wolstencroft E. C., Lorson C., Bassell G. J., Burghes A. H. M. and Beattie C. E. (2006) Survival motor neuron function in motor axons is independent of functions required for small nuclear ribonucleoprotein biogenesis. *J. Neurosci.* **26**, 11014–11022.

Cauchi R. J. (2014) Gem depletion: amyotrophic lateral sclerosis and spinal muscular atrophy crossover. *CNS Neurosci. Ther.* **20**, 574–581.

Chari A., Paknia E. and Fischer U. (2009) The role of RNP biogenesis in spinal muscular atrophy. *Curr. Opin. Cell Biol.* **21**, 387–393.

Chen Y.-Z., Bennett C. L., Huynh H. M. *et al.* (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.* **74**, 1128–1135.

Chen C., Nott T. J., Jin J. and Pawson T. (2011) Deciphering arginine methylation: Tudor tells the tale. *Nat. Rev. Mol. Cell Biol.* **12**, 629–642.

Chen S., Sayana P., Zhang X. and Le W. (2013) Genetics of amyotrophic lateral sclerosis: an update. *Mol. Neurodegener.* **8**, 28.

Cioce M. and Lamond A. I. (2005) Cajal bodies: a long history of discovery. *Annu. Rev. Cell Dev. Biol.* **21**, 105–131.

Coelho M. B., Attig J., Bellora N., König J., Hallegger M., Kayikci M., Eyras E., Ule J. and Smith C. W. J. (2015) Nuclear matrix protein Matrin3 regulates alternative splicing and forms overlapping regulatory networks with PTB. *EMBO J.* **34**, 653–668.

Cooper T. A., Wan L. and Dreyfuss G. (2009) *RNA and disease. Cell* **136**, 777–793.

Cooper-Knock J., Kirby J., Ferraiuolo L., Heath P. R., Rattray M. and Shaw P. J. (2012) Gene expression profiling in human neurodegenerative disease. *Nat. Rev. Neurol.* **8**, 518–530.

Corcia P., Ingre C., Blasco H. *et al.* (2012) Homozygous SMN2 deletion is a protective factor in the Swedish ALS population. *Eur. J. Hum. Genet.* **20**, 588–591.

DeJesus-Hernandez M., Mackenzie I. R., Boeve B. F. *et al.* (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256.

Fallini C., Bassell G. J. and Rossoll W. (2012a) Spinal muscular atrophy: the role of SMN in axonal mRNA regulation. *Brain Res.* **1462**, 81–92.

Fallini C., Bassell G. J. and Rossoll W. (2012b) The ALS disease protein TDP-43 is actively transported in motor neuron axons and regulates axon outgrowth. *Hum. Mol. Genet.* **21**, 3703–3718.

Feng W., Gubitz A. K., Wan L., Battle D. J., Dostie J., Golembe T. J. and Dreyfuss G. (2005) Gemins modulate the expression and activity of the SMN complex. *Hum. Mol. Genet.* **14**, 1605–1611.

Fox-Walsh K. L. and Hertel K. J. (2009) Splice-site pairing is an intrinsically high fidelity process. *Proc. Natl Acad. Sci. USA* **106**, 1766–1771.

Freibaum B. D., Lu Y., Lopez-Gonzalez R., Kim N. C. *et al.* (2015) GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature* **525**, 129–133.

Frey M. R. and Matera A. G. (1995) Coiled bodies contain U7 small nuclear RNA and associate with specific DNA sequences in interphase human cells. *Proc. Natl Acad. Sci.* **92**, 5915–5919.

Furger A., O'Sullivan J. M., Binnie A., Lee B. A. and Proudfoot N. J. (2002) Promoter proximal splice sites enhance transcription. *Genes Dev.* **16**, 2792–2799.

Gall J. G. (2000) Cajal bodies: the first 100 years. *Annu. Rev. Cell Dev. Biol.* **16**, 273–300.

Garcia E. L., Lu Z., Meers M. P., Praveen K. and Matera A. G. (2013) Developmental arrest of *Drosophila* survival motor neuron (Smn) mutants accounts for differences in expression of minor intron-containing genes. *RNA* **19**, 1510–1516.

Garcia E. L., Wen Y., Praveen K. and Matera A. G. (2016) Transcriptomic comparison of *Drosophila* snRNP biogenesis mutants reveals mutant-specific changes in pre-mRNA processing: implications for spinal muscular atrophy. *RNA* **22**, 1215–1227.

Gerbino V., Carrì M. T., Cozzolino M. and Achsel T. (2013) Mislocalised FUS mutants stall spliceosomal snRNPs in the cytoplasm. *Neurobiol. Dis.* **55**, 120–128.

Ginno P. A., Lott P. L., Christensen H. C., Korf I. and Chédin F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825.

Gitler A. D. and Tsuiji H. (2016) There has been an awakening: emerging mechanisms of C9orf72 mutations in FTD/ALS. *Brain Res.* **1647**, 19–29.

Greenway M. J., Alexander M. D., Ennis S., Traynor B. J., Corr B., Frost E., Green A. and Hardiman O. (2004) A novel candidate region for ALS on chromosome 14q11.2. *Neurology* **63**, 1936–1938.

Groen E. J. N., Fumoto K., Blokhuis A. M. *et al.* (2013) ALS-associated mutations in FUS disrupt the axonal distribution and function of SMN. *Hum. Mol. Genet.* **22**, 3690–3704.

Grohmann K., Schuelke M., Diers A. *et al.* (2001) Mutations in the gene encoding immunoglobulin mu-binding protein 2 cause spinal muscular atrophy with respiratory distress type 1. *Nat. Genet.* **29**, 75–77.

Gunderson S. I., Polycarpou-Schwarz M. and Mattaj I. W. (1998) U1 snRNP inhibits Pre-mRNA polyadenylation through a direct interaction between U1 70K and Poly(A) polymerase. *Mol. Cell* **1**, 255–264.

Haass C. and Selkoe D. J. (2007) Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **8**, 101–112.

Hashemi S. H., Anderson S. K., Huang Y., Lynch D. R. and Glass I. A. (2006) Senataxin, the yeast Sen1p orthologue: characterization of a unique protein in which recessive mutations cause ataxia and dominant mutations cause motor neuron disease. *Neurobiol. Dis.* **23**, 97–108.

He W. and Parker R. (2000) Functions of Lsm proteins in mRNA degradation and splicing. *Curr. Opin. Cell Biol.* **12**, 346–350.

Hein M. Y., Hubner N. C., Poser I. *et al.* (2015) A human interactome in three quantitative dimensions organized by stoichiometries and abundances. *Cell* **163**, 712–723.

Honda D., Ishigaki S., Iguchi Y., Fujioka Y., Udagawa T., Masuda A., Ohno K., Katsuno M. and Sobue G. (2014) The ALS/FTLD-related RNA-binding proteins TDP-43 and FUS have common downstream RNA targets in cortical neurons. *FEBS Open Bio* **4**, 1–10.

Iguchi Y., Katsuno M., Niwa J.-I. *et al.* (2013) Loss of TDP-43 causes age-dependent progressive motor neuron degeneration. *Brain* **136**, 1371–1382.

Ishihara T., Ariizumi Y., Shiga A. *et al.* (2013) Decreased number of Gemini of coiled bodies and U12 snRNA level in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **22**, 4136–4147.

Jobert L., Pinzón N., Van Herreweghe E., Jády B. E., Guialis A., Kiss T. and Tora L. (2009) Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep.* **10**, 494–500.

Johnson J. O., Pioro E. P., Boehringer A. *et al.* (2014) Mutations in the Matrin 3 gene cause familial amyotrophic lateral sclerosis. *Nat. Neurosci.* **17**, 664–666.

Jones K. W., Gorzynski K., Hales C. M., Fischer U., Badbanchi F., Terns R. M. and Terns M. P. (2001) Direct interaction of the spinal muscular atrophy disease protein SMN with the small nucleolar RNA-associated protein fibrillarin. *The Journal of Biological Chemistry* **276**, 38645–38651.

Jovičić A., Mertens J., Boeynaems S. *et al.* (2015) Modifiers of C9orf72 dipeptide repeat toxicity connect nucleocytoplasmic transport defects to FTD/ALS. *Nat. Neurosci.* **18**, 1226–1229.

Kaida D., Berg M. G., Younis I., Kasim M., Singh L. N., Wan L. and Dreyfuss G. (2010) U1 snRNP protects pre-mRNAs from premature cleavage and polyadenylation. *Nature* **468**, 664–668.

Kanekura K., Suzuki H., Aiso S. and Matsuoka M. (2009) ER stress and unfolded protein response in amyotrophic lateral sclerosis. *Mol. Neurobiol.* **39**, 81–89.

Kariya S., Re D. B., Jacquier A., Nelson K., Przedborski S. and Monani U. R. (2012) Mutant superoxide dismutase 1 (SOD1), a cause of amyotrophic lateral sclerosis, disrupts the recruitment of SMN, the spinal muscular atrophy protein to nuclear Cajal bodies. *Hum. Mol. Genet.* **21**, 3421–3434.

Kim H. J., Kim N. C., Wang Y.-D. *et al.* (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**, 467–473.

Kotova E., Jarnik M. and Tulin A. V. (2009) Poly (ADP-ribose) polymerase 1 is required for protein localization to Cajal body. *PLoS Genet.* **5**, e1000387.

Kwek K. Y., Murphy S., Furger A., Thomas B., O'Gorman W., Kimura H., Proudfoot N. J. and Akoulitchev A. (2002) U1 snRNA associates with TFIIH and regulates transcriptional initiation. *Nat. Struct. Biol.* **9**, 800–805.

Kwiatkowski T. J., Bosco D. A., Leclerc A. L. *et al.* (2009) Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323**, 1205–1208.

Kwon I., Xiang S., Kato M., Wu L., Theodoropoulos P., Wang T., Kim J., Yun J., Xie Y. and McKnight S. L. (2014) Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145.

Lagier-Tourenne C., Polymenidou M. and Cleveland D. W. (2010) TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. *Hum. Mol. Genet.* **19**, R46–R64.

Lagier-Tourenne C., Polymenidou M., Hutt K. R. *et al.* (2012) Divergent roles of ALS-linked proteins FUS/TLS and TDP-43 intersect in processing long pre-mRNAs. *Nat. Neurosci.* **15**, 1488–1497.

Lee Y.-B., Chen H.-J., Peres J. N. *et al.* (2013) Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep.* **5**, 1178–1186.

Lefebvre S., Bürglen L., Reboullet S., Clermont O., Burlet P., Viollet L., Benichou B., Cruaud C., Millasseau P. and Zeviani M. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* **80**, 155–165.

Li D. K., Tisdale S., Lotti F. and Pellizzoni L. (2014) SMN control of RNP assembly: from post-transcriptional gene regulation to motor neuron disease. *Semin. Cell Dev. Biol.* **32**, 22–29.

Ling S.-C., Albuquerque C. P., Han J. S., Lagier-Tourenne C., Tokunaga S., Zhou H. and Cleveland D. W. (2010) ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. *Proc. Natl Acad. Sci. USA* **107**, 13318–13323.

Ling J. P., Pletnikova O., Troncoso J. C. and Wong P. C. (2015) TDP-43 repression of nonconserved cryptic exons is compromised in ALS-FTD. *Science* **349**, 650–655.

Liu Q. and Dreyfuss G. (1996) A novel nuclear structure containing the survival of motor neurons protein. *EMBO J.* **15**, 3555–3565.

Liu Q., Fischer U., Wang F. and Dreyfuss G. (1997) The spinal muscular atrophy disease gene product, SMN, and its associated protein SIP1 are in a complex with spliceosomal snRNP proteins. *Cell* **90**, 1013–1021.

Liu Y., Yu J.-T., Zong Y., Zhou J. and Tan L. (2014) C9ORF72 mutations in neurodegenerative diseases. *Mol. Neurobiol.* **49**, 386–398.

Lorson C. L., Hahnen E., Androphy E. J. and Wirth B. (1999) A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc. Natl Acad. Sci.* **96**, 6307–6311.

Lorson C. L., Rindt H. and Shababi M. (2010) Spinal muscular atrophy: mechanisms and therapeutic strategies. *Hum. Mol. Genet.* **19**, R111–R118.

Lotti F., Imlach W. L., Saieva L. *et al.* (2012) An SMN-dependent U12 splicing event essential for motor circuit function. *Cell* **151**, 440–454.

Machyna M., Heyn P. and Neugebauer K. M. (2013) Cajal bodies: where form meets function. *Wiley Interdiscip Rev RNA* **4**, 17–34.

Mackenzie I. R. A. and Neumann M. (2016) Molecular neuropathology of frontotemporal dementia: insights into disease mechanisms from postmortem studies. *J. Neurochem.* **138**(Suppl 1), 54–70.

Mackenzie I. R. A., Bigio E. H., Ince P. G. *et al.* (2007) Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann. Neurol.* **61**, 427–434.

Mackenzie I. R., Rademakers R. and Neumann M. (2010) TDP-43 and FUS in amyotrophic lateral sclerosis and frontotemporal dementia. *Lancet Neurol* **9**, 995–1007.

Marzluff W. F., Wagner E. J. and Duronio R. J. (2008) Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat. Rev. Genet.* **9**, 843–854.

Masuda A., Takeda J.-I., Okuno T., Okamoto T., Ohkawara B., Ito M., Ishigaki S., Sobue G. and Ohno K. (2015) Position-specific binding of FUS to nascent RNA regulates mRNA length. *Genes Dev.* **29**, 1045–1057.

Matera A. G. and Wang Z. (2014) A day in the life of the spliceosome. *Nat. Rev. Mol. Cell Biol.* **15**, 108–121.

Menzies F. M., Fleming A. and Rubinsztein D. C. (2015) Compromised autophagy and neurodegenerative diseases. *Nat. Rev. Neurosci.* **16**, 345–357.

Monani U. R. (2005) Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease. *Neuron* **48**, 885–896.

Monani U. R., Lorson C. L., Parsons D. W., Prior T. W., Androphy E. J., Burghes A. H. and McPherson J. D. (1999) A single nucleotide difference that alters splicing patterns distinguishes the SMA gene SMN1 from the copy gene SMN2. *Hum. Mol. Genet.* **8**, 1177–1183.

Moreira M.-C., Klur S., Watanabe M., Németh A. H. *et al.* (2004) Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2. *Nat. Genet.* **36**, 225–227.

Nakaya T., Alexiou P., Maragkakis M., Chang A. and Mourelatos Z. (2013) FUS regulates genes coding for RNA-binding proteins in neurons by binding to their highly conserved introns. *RNA* **19**, 498–509.

Neumann M., Sampathu D. M., Kwong L. K. *et al.* (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133.

Nizami Z., Deryusheva S. and Gall J. G. (2010) The Cajal body and histone locus body. *Cold Spring Harb Perspect Biol* **2**, a000653.

Ou S. H., Wu F., Harrich D., García-Martínez L. F. and Gaynor R. B. (1995) Cloning and characterization of a novel cellular protein,

TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.* **69**, 3584–3596.

Patel A., Lee H. O., Jawerth L. *et al.* (2015) A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066–1077.

Peeters K., Chamova T. and Jordanova A. (2014) Clinical and genetic diversity of SMN1-negative proximal spinal muscular atrophies. *Brain* **137**, 2879–2896.

Pellizzoni L., Charroux B., Rappsilber J., Mann M. and Dreyfuss G. (2001) A functional interaction between the survival motor neuron complex and RNA polymerase II. *J. Cell Biol.* **152**, 75–85.

Perera N. D., Sheean R. K., Crouch P. J., White A. R., Horne M. K. and Turner B. J. (2016) Enhancing survival motor neuron expression extends lifespan and attenuates neurodegeneration in mutant TDP-43 mice. *Hum. Mol. Genet.* ddw247.

Polymenidou M., Lagier-Tourenne C., Hutt K. R. *et al.* (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. *Nat. Neurosci.* **14**, 459–468.

Porrua O. and Libri D. (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat. Rev. Mol. Cell Biol.* **16**, 190–202.

Praveen K., Wen Y. and Matera A. G. (2012) A *Drosophila* model of spinal muscular atrophy uncouples snRNP biogenesis functions of survival motor neuron from locomotion and viability defects. *Cell Rep.* **1**, 624–631.

Raczynska K. D., Ruepp M.-D., Brzek A., Reber S., Romeo V., Rindlisbacher B., Heller M., Szweykowska-Kulinska Z., Jarmolowski A. and Schümperli D. (2015) FUS/TLS contributes to replication-dependent histone gene expression by interaction with U7 snRNPs and histone-specific transcription factors. *Nucleic Acids Res.* **43**, 9711–9728.

Renton A. E., Majounie E., Waite A. *et al.* (2011) A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268.

Renton A. E., Chiò A. and Traynor B. J. (2014) State of play in amyotrophic lateral sclerosis genetics. *Nat. Neurosci.* **17**, 17–23.

Robberecht W. and Philips T. (2013) The changing scene of amyotrophic lateral sclerosis. *Nat. Rev. Neurosci.* **14**, 248–264.

Rogelj B., Easton L. E., Bogu G. K. *et al.* (2012) Widespread binding of FUS along nascent RNA regulates alternative splicing in the brain. *Sci. Rep.* **2**, 603.

Rossoll W. and Bassell G. J. (2009) Spinal Muscular Atrophy and a Model for Survival of Motor Neuron Protein Function in Axonal Ribonucleoprotein Complexes, *Cell Biology of the Axon*, vol. **48**, pp. 87–107. Springer, Berlin Heidelberg, Berlin, Heidelberg.

Rudnik-Schöneborn S., Arning L., Epplen J. T. and Zerres K. (2012) SETX gene mutation in a family diagnosed autosomal dominant proximal spinal muscular atrophy. *Neuromuscul. Disord.* **22**, 258–262.

Rulten S. L., Rotheray A., Green R. L., Grundy G. J., Moore D. A. Q., Gómez-Herreros F., Hafezparast M. and Caldecott K. W. (2014) PARP-1 dependent recruitment of the amyotrophic lateral sclerosis-associated protein FUS/TLS to sites of oxidative DNA damage. *Nucleic Acids Res.* **42**, 307–314.

Rutherford N. J., Zhang Y.-J., Baker M. *et al.* (2008) Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS Genet.* **4**, e1000193.

Sabra M., Texier P., El Maalouf J. and Lomonte P. (2013) The tudor protein survival motor neuron (SMN) is a chromatin-binding protein that interacts with methylated histone H3 lysine 79. *J. Cell Sci.* **126**, 3664–3677.

Salton M., Elkon R., Borodina T., Davydov A., Yaspo M.-L., Halperin E. and Shiloh Y. (2011) Matrin 3 binds and stabilizes mRNA. *PLoS ONE* **6**, e23882.

Salvi J. S. and Mekhail K. (2015) R-loops highlight the nucleus in ALS. *Nucleus* **6**, 23–29.

Sama R. R. K., Ward C. L. and Bosco D. A. (2014) Functions of FUS/TLS from DNA repair to stress response: implications for ALS. *ASN Neuro* **6**, 1–8.

Santos-Pereira J. M. and Aguilera A. (2015) R loops: new modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597.

Saxena S., Cabuy E. and Caroni P. (2009) A role for motoneuron subtype-selective ER stress in disease manifestations of FALS mice. *Nat. Neurosci.* **12**, 627–636.

Schreiber V., Dantzer F., Ame J.-C. and de Murcia G. (2006) Poly(ADP-ribose): novel functions for an old molecule. *Nat. Rev. Mol. Cell Biol.* **7**, 517–528.

Senderek J., Garvey S. M., Krieger M. *et al.* (2009) Autosomal-dominant distal myopathy associated with a recurrent missense mutation in the gene encoding the nuclear matrix protein, matrin 3. *Am. J. Hum. Genet.* **84**, 511–518.

Shababi M., Lorson C. L. and Rudnik-Schöneborn S. S. (2013) Spinal muscular atrophy: a motor neuron disorder or a multi-organ disease? *J. Anat.* **224**, 15–28.

Simic G. (2008) Pathogenesis of proximal autosomal recessive spinal muscular atrophy. *Acta Neuropathol.* **116**, 223–234.

Singh R. and Valcárcel J. (2005) Building specificity with nonspecific RNA-binding proteins. *Nat. Struct. Mol. Biol.* **12**, 645–653.

Skourti-Stathaki K., Proudfoot N. J. and Gromak N. (2011) Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* **42**, 794–805.

So B. R., Wan L., Zhang Z., Li P., Babiash E., Duan J., Younis I. and Dreyfuss G. (2016) A U1 snRNP-specific assembly pathway reveals the SMN complex as a versatile hub for RNP exchange. *Nat. Struct. Mol. Biol.* **23**, 225–230.

Spiluttini B., Gu B., Belagal P., Smirnova A. S., Nguyen V. T., Hébert C., Schmidt U., Bertrand E., Darzacq X. and Bensaude O. (2010) Splicing-independent recruitment of U1 snRNP to a transcription unit in living cells. *J. Cell Sci.* **123**, 2085–2093.

Sreedharan J., Blair I. P., Tripathi V. B. *et al.* (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319**, 1668–1672.

Stevens B., Allen N. J., Vazquez L. E. *et al.* (2007) The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164–1178.

Sugarman E. A., Nagan N., Zhu H. *et al.* (2012) Pan-ethnic carrier screening and prenatal diagnosis for spinal muscular atrophy: clinical laboratory analysis of >72,400 specimens. *Eur. J. Hum. Genet.* **20**, 27–32.

Sun S., Ling S.-C., Qiu J. *et al.* (2015) ALS-causative mutations in FUS/TLS confer gain and loss of function by altered association with SMN and U1-snRNP. *Nat. Commun.* **6**, 6171.

Suraweera A., Becherel O. J., Chen P., Rundle N., Woods R., Nakamura J., Gatei M. *et al.* (2007) Senataxin, defective in ataxia oculomotor apraxia type 2, is involved in the defense against oxidative DNA damage. *J. Cell Biol.* **177**, 969–979.

Suraweera A., Lim Y., Woods R., Birrell G. W., Nasim T., Becherel O. J. and Lavin M. F. (2009) Functional role for senataxin, defective in ataxia oculomotor apraxia type 2, in transcriptional regulation. *Hum. Mol. Genet.* **18**, 3384–3396.

Swinnen B. and Robberecht W. (2014) The phenotypic variability of amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **10**, 661–670.

Takahama K., Arai S., Kurokawa R. and Oyoshi T. (2009) Identification of DNA binding specificity for TLS. *Nucleic Acids Symp. Ser.* **53**, 247–248.

Tan A. Y. and Manley J. L. (2009) *The TET Family of Proteins: Functions and Roles in Disease.* **1**, 82–92.

Taylor J. E., Thomas N. H., Lewis C. M., Abbs S. J., Rodrigues N. R., Davies K. E. and Mathew C. G. (1998) Correlation of SMNt and SMNc gene copy number with age of onset and survival in spinal muscular atrophy. *Eur. J. Hum. Genet.* **6**, 467–474.

Tisdale S., Lotti F., Saieva L., Van Meerbeke J. P., Crawford T. O., Sumner C. J., Mentis G. Z. and Pellizzoni L. (2013) SMN is essential for the biogenesis of U7 small nuclear ribonucleoprotein and 3′-end formation of histone mRNAs. *Cell Rep.* **5**, 1187–1195.

Tsuiji H., Iguchi Y., Furuya A. *et al.* (2013) Spliceosome integrity is defective in the motor neuron diseases ALS and SMA. *EMBO Mol. Med.* **5**, 221–234.

Turner B. J., Alfazema N., Sheean R. K., Sleigh J. N., Davies K. E., Horne M. K. and Talbot K. (2014) Overexpression of survival motor neuron improves neuromuscular function and motor neuron survival in mutant SOD1 mice. *Neurobiol. Aging* **35**, 906–915.

Vance C., Rogelj B., Hortobágyi T., De Vos K. J., Nishimura A. L., Sreedharan J., Hu X. *et al.* (2009) Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323**, 1208–1211.

Vance C., Scotter E. L., Nishimura A. L. *et al.* (2013) ALS mutant FUS disrupts nuclear localization and sequesters wild-type FUS within cytoplasmic stress granules. *Hum. Mol. Genet.* **22**, 2676–2688.

Vanden Broeck L., Callaerts P. and Dermaut B. (2014) TDP-43-mediated neurodegeneration: towards a loss-of-function hypothesis? *Trends Mol. Med.* **20**, 66–71.

Vantaggiato C., Bondioni S., Airoldi G., Bozzato A., Borsani G., Rugarli E. I., Bresolin N., Clementi E. and Bassi M. T. (2011) Senataxin modulates neurite growth through fibroblast growth factor 8 signalling. *Brain* **134**, 1808–1828.

Veldink J. H., Kalmijn S., Van der Hout A. H., Lemmink H. H., Groeneveld G. J., Lummen C., Scheffer H., Wokke J. H. J. and Van den Berg L. H. (2005) SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. *Neurology* **65**, 820–825.

Wang Q., Sawyer I. A., Sung M.-H., Sturgill D., Shevtsov S. P., Pegoraro G., Hakim O., Baek S., Hager G. L. and Dundr M. (2016) Cajal bodies are linked to genome conformation. *Nat. Commun.* **7**, 10966.

West S. (2012) The increasing functional repertoire of U1 snRNA. *Biochem. Soc. Trans.* **40**, 846–849.

Winkler C., Eggert C., Gradl D., Meister G., Giegerich M., Wedlich D., Laggerbauer B. and Fischer U. (2005) Reduced U snRNP assembly causes motor axon degeneration in an animal model for spinal muscular atrophy. *Genes Dev.* **19**, 2320–2330.

Workman E., Saieva L., Carrel T. L., Crawford T. O., Liu D., Lutz C., Beattie C. E., Pellizzoni L. and Burghes A. H. M. (2009) A SMN missense mutation complements SMN2 restoring snRNPs and rescuing SMA mice. *Hum. Mol. Genet.* **18**, 2215–2229.

Yamazaki T., Chen S., Yu Y. *et al.* (2012) FUS-SMN protein interactions link the motor neuron diseases ALS and SMA. *Cell Rep.* **2**, 799–806.

Yong J., Kasim M., Bachorik J. L., Wan L. and Dreyfuss G. (2010) Gemin5 delivers snRNA precursors to the SMN complex for snRNP biogenesis. *Mol. Cell* **38**, 551–562.

Yu Y. and Reed R. (2015) FUS functions in coupling transcription to splicing by mediating an interaction between RNAP II and U1 snRNP. *Proc. Natl Acad. Sci. USA* **112**, 8608–8613.

Zhang Z., Lotti F., Dittmar K., Younis I., Wan L., Kasim M. and Dreyfuss G. (2008) SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* **133**, 585–600.

Zhang Z., Pinto A. M., Wan L., Wang W., Berg M. G., Oliva I., Singh L. N., Dengler C., Wei Z. and Dreyfuss G. (2013) Dysregulation of synaptogenesis genes antecedes motor neuron pathology in spinal muscular atrophy. *Proc. Natl Acad. Sci. USA* **110**, 19348–19353.

Zhang K., Donnelly C. J., Haeusler A. R. *et al.* (2015) The C9orf72 repeat expansion disrupts nucleocytoplasmic transport. *Nature* **525**, 56–61.

Zhao D. Y., Gish G., Braunschweig U., Li Y., Ni Z., Schmitges F. W., Zhong G. *et al.* (2016) SMN and symmetric arginine dimethylation of RNA polymerase II C-terminal domain control termination. *Nature* **529**, 48–53.

Zheng S. and Black D. L. (2013) Alternative pre-mRNA splicing in neurons: growing up and extending its reach. *Trends in Genetics : TIG* **29**, 442–448.