



Targeted Genetic Screen in Amyotrophic Lateral Sclerosis Reveals Novel Genetic Variants with Synergistic Effect on Clinical Phenotype

Johnathan Cooper-Knock, Henry Robins, Isabell Niedermoser, Matthew Wyles, Paul R. Heath, Adrian Higginbottom, Theresa Walsh, Mbombe Kazoka, Project MinE ALS Sequencing Consortium[†], Paul G. Ince, Guillaume M. Hautbergue, Christopher J. McDermott, Janine Kirby and Pamela J. Shaw*

Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, United Kingdom

OPEN ACCESS

Edited by:

Michael J. Schmeisser,
Medizinische Fakultät,
Universitätsklinikum Magdeburg,
Germany

Reviewed by:

Luc Dupuis,
INSERM U1118, France
Francesco Roselli,
University of Ulm, Germany

*Correspondence:

Pamela J. Shaw
pamela.shaw@sheffield.ac.uk

[†]The author names are listed in the section "Project MinE ALS Sequencing Consortium."

Received: 18 August 2017

Accepted: 26 October 2017

Published: 09 November 2017

Citation:

Cooper-Knock J, Robins H, Niedermoser I, Wyles M, Heath PR, Higginbottom A, Walsh T, Kazoka M, Project MinE ALS Sequencing Consortium, Ince PG, Hautbergue GM, McDermott CJ, Kirby J and Shaw PJ (2017) Targeted Genetic Screen in Amyotrophic Lateral Sclerosis Reveals Novel Genetic Variants with Synergistic Effect on Clinical Phenotype. *Front. Mol. Neurosci.* 10:370. doi: 10.3389/fnmol.2017.00370

Amyotrophic lateral sclerosis (ALS) is underpinned by an oligogenic rare variant architecture. Identified genetic variants of ALS include RNA-binding proteins containing prion-like domains (PrLDs). We hypothesized that screening genes encoding additional similar proteins will yield novel genetic causes of ALS. The most common genetic variant of ALS patients is a G4C2-repeat expansion within *C9ORF72*. We have shown that G4C2-repeat RNA sequesters RNA-binding proteins. A logical consequence of this is that loss-of-function mutations in G4C2-binding partners might contribute to ALS pathogenesis independently of and/or synergistically with *C9ORF72* expansions. Targeted sequencing of genomic DNA encoding either RNA-binding proteins or known ALS genes ($n = 274$ genes) was performed in ALS patients to identify rare deleterious genetic variants and explore genotype-phenotype relationships. Genomic DNA was extracted from 103 ALS patients including 42 familial ALS patients and 61 young-onset (average age of onset 41 years) sporadic ALS patients; patients were chosen to maximize the probability of identifying genetic causes of ALS. Thirteen patients carried a G4C2-repeat expansion of *C9ORF72*. We identified 42 patients with rare deleterious variants; 6 patients carried more than one variant. Twelve mutations were discovered in known ALS genes which served as a validation of our strategy. Rare deleterious variants in RNA-binding proteins were significantly enriched in ALS patients compared to control frequencies ($p = 5.31E-18$). Nineteen patients featured at least one variant in a RNA-binding protein containing a PrLD. The number of variants per patient correlated with rate of disease progression (t -test, $p = 0.033$). We identified eighteen patients with a single variant in a G4C2-repeat binding protein. Patients with a G4C2-binding protein variant in combination with a *C9ORF72* expansion had a significantly faster disease course (t -test, $p = 0.025$). Our data are consistent with an oligogenic model of ALS. We provide evidence for a number of entirely novel genetic variants of ALS caused by mutations in

RNA-binding proteins. Moreover we show that these mutations act synergistically with each other and with *C9ORF72* expansions to modify the clinical phenotype of ALS. A key finding is that this synergy is present only between functionally interacting variants. This work has significant implications for ALS therapy development.

Keywords: amyotrophic lateral sclerosis, RNA binding proteins, oligogenic inheritance, *C9ORF72*, DNA sequencing

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is an age-related neurodegenerative disorder. The lifetime risk of ALS is ~1 in 400. The ALS phenotype is markedly variable but ~80% of patients die from respiratory failure within 2–5 years (Cooper-Knock et al., 2013). The majority of ALS is apparently sporadic, but 5–10% of patients show autosomal dominant inheritance. It is recognized that ALS is likely to be an oligogenic disorder even when it is apparently sporadic (van Blitterswijk et al., 2012). A mixed-model association analysis in 12,577 ALS cases and 23,475 controls was consistent with an oligogenic rare variant architecture (van Rheen et al., 2016).

Identified ALS loci highlight a small number of pathways, most prominent of which is RNA metabolism. Pathogenic mutations have been discovered in multiple RNA-recognition motif (RRM) containing proteins including EWSR1, FUS, HNRNPA1, HNRNPA2B1, TAF15, and TDP-43 (Cooper-Knock et al., 2012). All of these proteins contain prion-like domains (PrLDs) (Harrison and Shorter, 2017). A PrLD consists of low complexity sequence with an “infectious” conformation that allows these proteins to undergo liquid-phase transition. Physiologically, such transitions allow the formation of membrane-less organelles such as stress granules, but pathologically they are thought to lead to irreversible protein aggregation. Often membrane-less organelles contain RNA; in addition to PrLD interaction it has been shown that RRM interaction with RNA is essential for integrity of so-called RNA granules (Molliex et al., 2015). The infectious aspect of PrLDs refers to the ability of aggregated protein to induce an aggregation–conformation in unaggregated protein, which is a proposed mechanism for ALS disease spread through the CNS (Ravits, 2014).

Thirty-one of the 213 identified RRM-containing proteins in the human proteome rank in the top 250 most prion-like (Alberti et al., 2009; Couthouis et al., 2011); this includes EWSR1, FUS, TAF15, and TDP-43 which are known to be mutated in ALS cases. We screened 147 additional genes encoding RRM-containing proteins with prion-like domains for mutations in ALS cases.

In the most common genetic variant of ALS, patients carry a G4C2-repeat expansion within intron 1 of *C9ORF72* (DeJesus-Hernandez et al., 2011; Renton et al., 2011). *C9ORF72*-ALS patients represent the full spectrum of sporadic ALS both clinically and pathologically (Cooper-Knock et al., 2012). The mechanism of pathogenesis in these cases is unknown. Three mechanisms have been proposed and to some extent demonstrated: (1) Haploinsufficiency related to disrupted expression of the *C9ORF72* protein. (2) Gain-of-function toxicity

of G4C2-repeat RNA molecules transcribed from the mutated sequence. (3) Toxicity of dipeptide-repeat proteins translated from the repetitive RNA (Cooper-Knock et al., 2015b). It is hypothesized that G4C2-repeat RNA sequesters RNA-binding proteins away from their normal location causing a functional haploinsufficiency (Cooper-Knock et al., 2014). Notably the antisense transcript consisting of C4G2-repeat RNA binds a similar set of RNA-binding proteins (Cooper-Knock et al., 2015a). A logical consequence of this hypothesis is that loss-of-function mutations in G4C2-binding partners might contribute to ALS pathogenesis independently of and/or act synergistically with *C9ORF72* expansions. Evidence in myotonic dystrophy supports this hypothesis: mutations in muscleblind-like proteins modify the phenotype caused by sequestration of the same proteins by CUG-repeat RNA (e.g., Choi et al., 2016). Similarly mice lacking muscleblind-like 1 exhibit some of the features of myotonic dystrophy despite the absence of CUG-repeat RNA (Dixon et al., 2015).

We tested whether mutations in RNA-binding proteins, including both RRM-containing proteins with a PrLD and G4C2-binding partners, are a cause of ALS and/or whether they modify the clinical phenotype. Our patient cohort (Table 1, Supplementary Table 2) was comprised of either familial ALS cases caused by a *C9ORF72* expansion ($n = 13$) or FALS without a known genetic cause identified ($n = 42$) or young patients with sporadic ALS ($n = 61$) who are more likely to carry a pathogenic mutation than older patients with sporadic ALS (Cooper-Knock et al., 2013). Our filtering strategy aimed to identify rare deleterious variants rather than common low-risk variants. We also screened for variants in known ALS genes to augment the analysis and validate our strategy.

We identified a number of apparently toxic variants in RNA-binding proteins in ALS patients at a significantly higher frequency than is observed in normal controls. Moreover we showed that these variants act synergistically with each other and with known ALS-causing mutations to determine the clinical severity of ALS. This has important implications for future ALS-therapy development.

MATERIALS AND METHODS

Design of the Targeted Genetic Screen

The complete list of sequenced genes is provided in Supplementary Table 1. Genes were either known ALS genes or genes encoding RNA-binding proteins. The RNA-binding proteins were in two groups—RRM-containing proteins with a PrLD (Couthouis et al., 2011) or those identified binding

TABLE 1 | Summary of targeted DNA sequencing screen.

Group	Number of patients	Number with a newly identified variant	Number with > 1 newly identified variant	Average age of onset (standard deviation) (years)	Male:Female Ratio
Familial ALS	42	16	1	60 (8.6)	1.5:1
Young sporadic ALS	61	26	5	41 (15.8)	1.9:1
Total	103	42	6	49 (15.2)	1.8:1

partners of the G4C2-repeat expansion (Cooper-Knock et al., 2014).

Selection of Patients for Screening

ALS patients were selected to increase the probability of discovering novel genetic variants—they either had a positive family history, or they were relatively young (<50 years old) at presentation or they carried an expansion of *C9ORF72*. Genomic DNA was extracted from 103 ALS patients from the North of England. The cohort included 34 familial ALS patients in whom a genetic cause had not been identified despite screening for ALS associated mutations in *SOD1*, *C9ORF72*, *TARDBP*, and *FUS*; 61 young-onset sporadic ALS patients; and thirteen *C9ORF72*-ALS patients (Table 1). A patient with an identified mutation in *FUS* was included as a positive control. G4C2-repeat expansions of *C9ORF72* expansions were identified by repeat-primed PCR as described previously (Cooper-Knock et al., 2012); all patients were screened for *C9ORF72* expansion prior to selection for the screen. The study was approved by the South Sheffield Research Ethics Committee and informed consent was obtained for all samples.

DNA Sequencing

Genomic DNA was enriched for selected RNA-binding proteins and known ALS genes using a custom designed Agilent SureSelect in solution kit. Sequencing was performed using an Illumina HiScan platform according to manufacturers instructions.

Rare deleterious mutations were defined by frequency within the Exome Aggregation Consortium data set of <1/10,000 control alleles (Lek et al., 2016), and a Phred-scaled Combined Annotation Dependent Depletion (CADD) score >10 (Kircher et al., 2014). Comparison of various pathogenicity prediction tools recently supported the sensitivity and specificity of CADD (Salgado et al., 2016). Given that we were focused on exonic changes with an effect on protein function, synonymous changes were excluded. We excluded any changes with a read depth <10 and validated by Sanger sequencing any changes with read depth 10–15 or a novel allele frequency less than one third the reference allele frequency (Supplementary Figure 1).

ExAC defines constrained genes based on an observed frequency of loss of function mutations which is much less than predicted by sequence specific mutation probabilities (Lek et al., 2016). A threshold for “constrained” is set as probability of a gene being loss of function intolerant (PLI) > 0.95.

RESULTS

Our aim was to identify genetic changes which may cause or contribute to ALS pathogenesis. Consistent with an oligogenic rare variant architecture of ALS (van Rheenen et al., 2016) we proposed that such changes are unlikely to be common in the background population, but may be present. We filtered sequencing data for rare deleterious variants defined as frequency within the ExAC data set of <1/10,000 control alleles (Lek et al., 2016), and a Phred-scaled CADD score >10 (change is within 10% most deleterious reference variants) (Kircher et al., 2014). All genetic changes with a low read depth were validated by Sanger sequencing (Supplementary Figure 1).

In 42 (of 103 screened) patients we identified a rare deleterious variant; six patients carried more than one variant. Thirteen *C9ORF72*-ALS patients were included in the screen; in eight we identified an additional rare deleterious variant (i.e., in addition to a G4C2-repeat expansion of *C9ORF72*) and in two patients we identified more than one additional variant. Average disease duration for patients in the screen was 66 months; average disease duration in patients with an identified variant was 61 months compared to 73 months in patients in whom no variant was identified, although this difference was not statistically significant (*t*-test, *p* = 0.14). In both patients with and without an identified variant average age of onset was 49 years.

Identified Mutations in Known ALS Genes

We identified 12 patients with mutations in nine known ALS genes (Table 2, Supplementary Table 2). This is expected based on reported frequencies of these mutations and served as a validation of our strategy. One patient with a previously identified *FUS* mutation was included as a positive control. Rare deleterious variants were newly identified in *ALS2*, *DCTN1* (two different variants), *ELP3*, *EWSR1*, *SETX* (two different variants), *SOD1* (two different variants), *UNC13A*, *C9ORF72*, and *VCP*.

Several of the mutations we identified in ALS genes affect previously reported amino acids or protein domains. For example, both mutations in *DCTN1* were within the dynein associated protein domain, which is consistent with previously reported mutations (Münch et al., 2004); the mutation identified in *EWSR1* occurs in the same amino acid as previously reported (Couthouis et al., 2011); one of the *SETX* mutations we identified lies within a helicase domain which contains several previously reported mutations (Hirano et al., 2011); both *SOD1* mutations have been previously described in familial ALS (Orrell et al., 1999); and a mutation in the same amino acid of *VCP* has been previously identified in another ALS patient (Johnson et al., 2010).

TABLE 2 | Identified rare deleterious variants in known ALS genes.

Gene	Mutation	Amino acid change	Mutated protein domain	Sporadic/Familial	CADD
<i>C9ORF72</i>	A1239G	I413M	Alpha domain	Sporadic	13.9
<i>DCTN1</i>	G1326A/G1668A/G1617A/G1707A/G1728A	M442I/M556I/M539I/M569I/M576I	Dynein associated protein domain	Sporadic	17.2
<i>DCTN1</i>	G1193C/G1535C/G1484C/G1574C/G1595C	R398P/R512P/R495P/R525P/R532P	Dynein associated protein domain	Sporadic	24.6
<i>ELP3</i>	T654A/T795A/T735A/T969A/T1101A	Y218X/Y265X/Y245X/Y323X/Y337X	Affects all functional domains	Familial	37
<i>EWSR1</i>	G1366A/G1531A/G1534A/G1549A	G456R/G511R/G512R/G517R	Within R/G/P-rich domain	Familial	18.3
<i>SETX</i>	A6172C	K2058Q	Helicase domain	Familial	12.8
<i>SETX</i>	C1750G	L584V	Outside described domains	Familial	13.4
<i>UNC13A</i>	G3091A	G1031R	Calcium dependent secretion activator domain	Familial	11
<i>SOD1</i>	G217A	G72S	Cu/Zn binding domain	Sporadic	36
<i>SOD1</i>	T341C	I113T	Cu/Zn binding domain	Sporadic	19.5
<i>ALS2</i>	G1681A	V561I	Regulator of chromatin condensation domain	Sporadic	14.5
<i>VCP</i>	G278A	R93H	Aspartate decarboxylase-like domain	Sporadic	21.8

Other variants we identified in known ALS genes are more novel. *ELP3* has been previously associated with ALS by GWAS (Simpson et al., 2009), but pathogenic variants have not been identified. The patient identified in this screen demonstrated a nonsense mutation in exon 10 which disrupts all described functional domains of the protein. Similarly variation in *UNC13A* has been identified as a risk factor for sporadic ALS (van Es et al., 2009) and as a modifier of the clinical phenotype, but pathogenic variants have not been identified. Our patient with a variant in *UNC13A* has a family history of ALS and no other identified mutation in an ALS gene (or any other gene in our screen). One sporadic ALS patient has a variant in *ALS2*; given that mutations in *ALS2* are usually autosomal recessive and associated with a slowly progressive juvenile onset form of the disease, then this variant is of unknown significance. However, no study has reported this exact change previously (Al-Chalabi et al., 2003; Luigetti et al., 2013). Similarly a rare deleterious variant was identified in *C9ORF72* in a patient who also carried a G4C2-repeat expansion; no pathogenic variants have been confirmed in *C9ORF72* except the G4C2-repeat expansion in intron 1 and therefore this variant is also of unknown significance.

Identified Rare Deleterious Variants in RRM-Containing Proteins with Prion-Like Domains

We identified 19 patients with a rare deleterious variant in a RNA-binding protein with a PrLD of whom three had more than one variant (Table 3, Supplementary Table 2). Fourteen of the patients had died, four patients were still alive and in these cases disease duration was censored to the present date. One patient with a variant in *MTHFSD* also had a mutation in *SOD1*. *SOD1* mutations are associated with a distinct clinical phenotype and pathology compared to characterized mutations in RNA-binding proteins (Cooper-Knock et al., 2013) and therefore this patient was excluded from further analysis. Of the 21 identified variants

remaining, 16 (76%) occurred in either the RRM domain or a low complexity sequence (Table 3). The number of variants per patient correlated with rate of disease progression (Figure 1A, *t*-test, $p = 0.033$) but not age of onset. Including *C9ORF72* expansions in this analysis did not appear to be synergistic.

The Project MinE browser (<http://databrowser.projectmine.com/>) was utilized to search for additional evidence of similar variants in these proteins. The Project MinE Consortium has to date reported whole genome sequencing of 1169 ALS cases and 608 controls from the Netherlands. For *RBM4B*, *RBM45*, *RBMS2*, *RAVER1*, *PPARGC1B*, and *TRNAU1AP*, the project Project MinE data identified additional variant(s) within the same exon which were present either exclusively in ALS patients or were more frequent in ALS patients than controls. *RBM12*, *RBM12B*, *RBM15*, *RBM15B*, and *RBM45* are single exon genes, but Project MinE identified ALS cases with disease-associated variant(s) within <25 amino acids in each of these genes. This clustering of cases for each of these genes supports the functional significance of the rare variants we have discovered.

The ALS Variant Server, Worcester, MA (<http://als.umassmed.edu/>) reports whole exome sequencing from 1,022 familial ALS patients. Within this cohort we identified an additional example of an ALS patient carrying an A622G variant in *RBM12* and four ALS patients carrying p.S550G/p.S525G/p.S589G (single case) or p.E395K/p.E370K/p.E434K (3 cases) variants in *PPARGC1B*.

It is noteworthy that a small number of genes found to contain rare deleterious variants but classified as known ALS genes or G4C2-repeat binding partners are also RRM-containing proteins with a PrLD. This includes *EWSR1*, *HNRNPA3*, *HNRNPU*, and *HNRNPUL1*. Except for being previously identified as a known ALS gene, *EWSR1* is not distinct from the other RRM-containing proteins with a PrLD under consideration; therefore *EWSR1* is included in the analysis of synergy detailed above. In contrast *HNRNPA3*, *HNRNPU*, and *HNRNPUL1* were selected on the basis of an independent hypothesis: that loss of function in the proteins encoded by these genes might mimic sequestration

TABLE 3 | Identified rare deleterious variants in RNA-binding proteins with prion-like domains.

Gene	Variant	Amino acid change	Mutated RRM/Low complexity domain	Sporadic/Familial	Exac constrained (PLi > 0.95)	CADD phred score	Additional variant
<i>NOL8</i>	T2597G/T2393G	L866R/L798R	E-rich domain	Sporadic	No	23.8	RBM4B
<i>RBM4B</i>	C701T	A234V	A-rich domain	Sporadic	No	15.2	NOL8
<i>EIF3B</i>	C943T	R315C	No	Familial	Yes	17.2	None
<i>RBM41</i>	G760A	A254T	No	Sporadic	No	11.3	None
<i>RBM12</i>	A622G	I208V	P-rich domain	Familial	No	12.4	RBM15
<i>RBM15</i>	G1787A	R596H	R-rich domain	Familial	Yes	20.6	RBM12
<i>HNRNPM</i>	G544A/G904A/G787A	G182S/G302S/G263S	No	Sporadic	Yes	22.7	None
<i>PPARGC1B</i>	C1037A/C962A/C1154A	P346H/P321H/P385H	No	Sporadic	No	12.1	None
<i>PPARGC1B</i>	A1648G/A1573G/A1765G	S550G/S525G/S589G	No	Sporadic	No	15.4	None
<i>PPARGC1B</i>	G1183A/G1108A/G1300A	E395K/E370K/E434K	E-rich domain	Sporadic	No	11.1	None
<i>MTHFSD</i>	G472C/G469C/G412C	A158P/A157P/A138P	No	Sporadic	No	22	None
<i>SPEN</i>	G1649A	R550H	RRM	Sporadic	Yes	34	None
<i>PABPC1L</i>	G808A	V270M	RRM	Sporadic	Yes	11.8	RBMXL3
<i>RBMXL3</i>	C362T	P121L	No	Sporadic	No	16.1	PABPC1L
<i>RAVER1</i>	T194G	L65R	RRM	Sporadic	Yes	28.3	None
<i>RBM12B</i>	A652T	M218L	RRM	Sporadic	No	17.9	None
<i>RBM15B</i>	G1385C	S462T	RRM	Sporadic	Yes	19.2	None
<i>RBM45</i>	G338A	R113Q	RRM	Sporadic	No	22.3	None
<i>RBMS2</i>	G354T	K118N	RRM	Familial	No	16.8	None
<i>RBMXL2</i>	G995T	R332L	R/E/P-rich domain	Sporadic	No	17.1	None
<i>TRNAU1AP</i>	G124T	G42W	RRM	Sporadic	No	29.6	None
<i>EWSR1</i>	G1366A/G1531A/ G1534A/G1549A	G456R/G511R/ G512R/G517R	Within R/G/P-rich domain	Familial	Yes	18.3	None

by G4C2-repeat-RNA derived from a *C9ORF72* expansion. The majority of variants identified in RRM-containing proteins with a PrLD are located in either the RRM-domain or the PrLD but, consistent with an alternate mechanism, variants identified in *HNRNPA3*, *HNRNPU*, and *HNRNPUL1* are located distinct functional domains (Table 4). To avoid potentially confounding discrepancy between mechanisms of pathogenicity *HNRNPA3*, *HNRNPU*, and *HNRNPUL1* were not included in analysis of other variants identified within RRM-containing proteins with a PrLD.

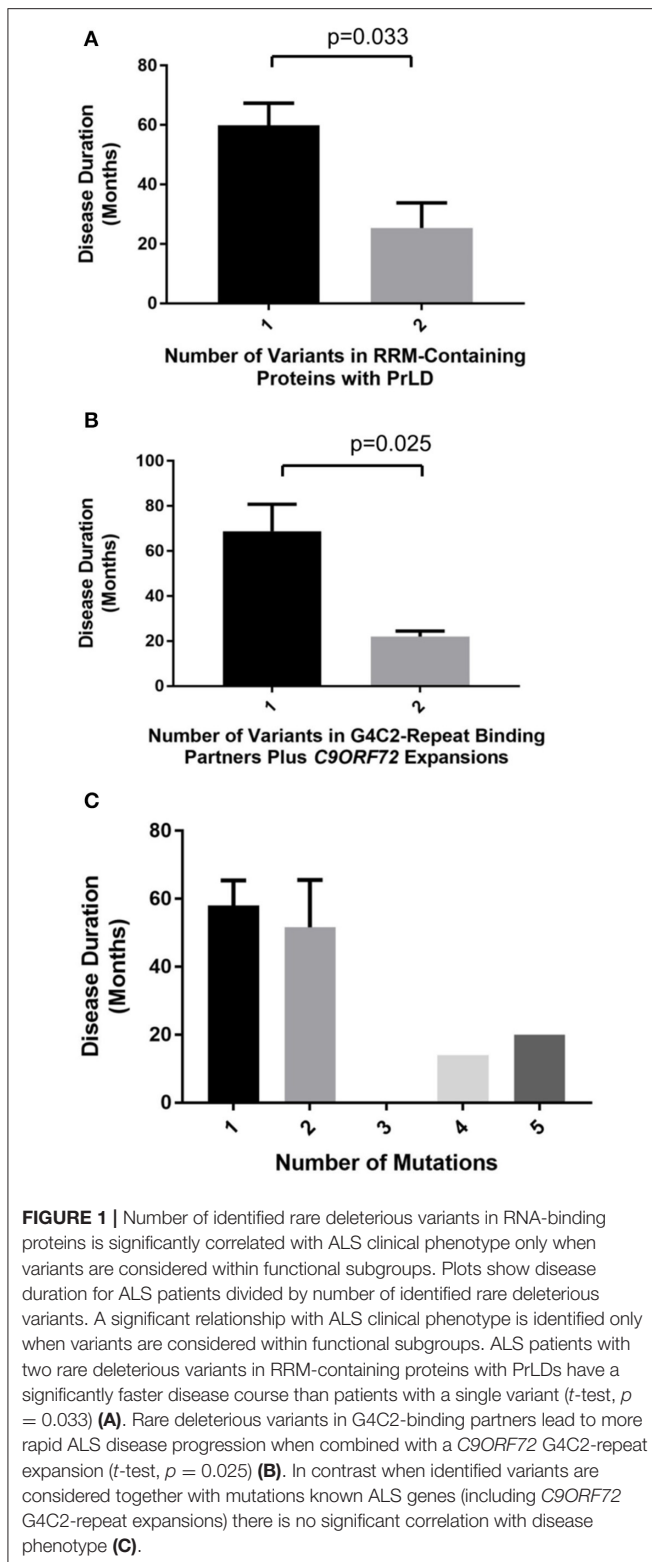
Identified Rare Deleterious Variants in G4C2-Repeat Binding Proteins

We identified 18 patients with a rare deleterious variant in a G4C2-repeat binding protein (Table 4, Supplementary Table 2). No patients had more than one variant in a G4C2-repeat binding protein. Five of the patients carried a G4C2-repeat expansion in *C9ORF72*. Fourteen of the patients had died, six patients were still alive and in these cases disease duration was censored to the present date. Patients with a G4C2-binding protein variant in combination with a *C9ORF72* expansion had a significantly faster disease course (Figure 1B, *t*-test, $p = 0.025$) but age of onset was not significantly different. For one patient with a variant in *ILF3*, no clinical information was available. In two specific examples the same gene is mutated in patients with and without a *C9ORF72* expansion—*SLC1A3* and *EEFIG*. In both cases there is a 50% reduction in disease duration (*SLC1A3*: 52 months to 27 months;

EEFIG: 79 months censored to 22 months) in the patient carrying the *C9ORF72* expansion and the mutation.

Sequestration of RNA-binding proteins by G4C2-repeat RNA associated with *C9ORF72*-ALS would be expected to prevent those proteins performing their normal function. Consequently a mutation which exacerbates this toxicity would be expected to cause loss-of-function. Of the 18 G4C2-repeat binding proteins in which we identified a rare deleterious variant, 67% are encoded by genes which are defined by ExAC as extremely loss-of-function intolerant (ExAC refer to this property as “constrained”) (Lek et al., 2016). This is enriched compared to the total list of G4C2-repeat binding proteins screened (Supplementary Table 1) of which 42% are ExAC constrained (41 constrained from 98 total). This observation supports our proposed mechanism. In comparison, for RRM-containing proteins with a PrLD, the proportion of variants discovered in ExAC constrained genes is only 32%.

The Project MinE browser (<http://databrowser.projectmine.com/>) was utilized to search for additional evidence of similar variants in these proteins. For *SLC1A3*, *EEFIG*, *hnRNPU*, *hnRNPU1*, *EZR*, and *GRSF1* Project MinE identified additional variant(s) within the same exon which were present either exclusively in ALS patients or were more frequent in ALS patients than controls. The ALS Variant Server, Worcester, MA (<http://als.umassmed.edu/>) reports whole exome sequencing from 1,022 familial ALS patients. Within this cohort we identified an additional example of an ALS patient carrying a p.D403E



mutation in *EEF1G*, a p.P12R mutation in *HNRNPA3*, and a p.A297P/p.A308P mutation in *SRPK2*. This clustering of cases for each of these genes supports the functional significance of the rare variants we have discovered.

Rare Deleterious Variants in RNA-Binding Proteins Are Enriched in ALS Cases

To calculate whether the observed frequency of rare deleterious variants in RNA-binding proteins in our DNA sequencing screen is higher than expected we utilized ExAC frequencies and CADD scores for the identified changes. CADD scoring is expressed as the observed frequency of variants which are at least as pathogenic as the observed variant. For this analysis we assumed that observed frequency is independent of pathogenicity on the basis that ALS does not usually affect reproductive fitness. We observed 39 rare deleterious variants in 1,223,647 bases of DNA from 103 patients; this is a significant enrichment compared to observed control frequencies ($p = 5.31E-18$) suggesting that these variants are significantly enriched in ALS patients.

Synergy between Variants Is Function Specific

No significant correlation was identified between total number of variants per patient and the clinical phenotype (Pearson correlation, correlation coefficient = -0.20 , $p = 0.21$) (Figure 1C). This was unchanged whether or not *C9ORF72* expansions are considered. In contrast, when either RRM-containing proteins with PrLDs or G4C2 binding partners are considered in isolation, then there is a significant synergistic effect on clinical phenotype (Figures 1A,B). We conclude that a synergy is present only between variants in functionally interacting genes/proteins.

DISCUSSION

A new period of ALS genetics has begun in which we need to think of ALS as not a predominantly sporadic disease with a small proportion of monogenic familial cases, but rather as a pathogenesis shaped by synergy between oligogenic rare variants. It is likely that many ALS-associated genetic variants do not cause disease except in combination with other genetic and environmental factors. This is consistent with ALS as a multistep process as proposed by Al-Chalabi et al. (2014). With an oligogenic model in mind, we performed targeted genetic sequencing of RNA-binding proteins in ALS patients and identified rare deleterious variants at a significantly higher than control frequency. We aimed to identify novel pathogenic mutations and to discover evidence that these mutations act synergistically to produce the ALS phenotype. We achieved this and for the first time we have shown that synergy between mutations is specific to groups of functionally related genes/proteins.

We have shown that rare deleterious variants in RRM-containing proteins with a PrLD act synergistically to determine speed of ALS progression. Synergy is consistent with action in a common pathway. PrLD are thought to facilitate protein-protein interactions which are key to the formation of membrane-less cellular compartments (March et al., 2016). Important examples of membrane-less compartments are RNA-protein complexes such as P-bodies and stress granules. These RNA granules

TABLE 4 | Identified rare deleterious variants in G4C2-repeat binding partners.

Gene	Variant	Amino acid change	Sporadic/Familial	Exac constrained (PLi > 0.95)	CADD	C9ORF72 expansion
<i>SLC1A3</i>	A509G/A647G	N170S/N216S	Sporadic	No	13.9	Yes
<i>SLC1A3</i>	C372G/C510G	F124L/F170L	Familial	No	10.1	No
<i>ATP5B</i>	T803C	V268A	Sporadic	Yes	25.5	No
<i>MYH9</i>	A3181T	S1061C	Sporadic	No	19.2	Yes
<i>EEF1G</i>	T1209A	D403E	Familial	Yes	14.5	Yes
<i>EEF1G</i>	C979T	R327C	Sporadic	Yes	18.3	No
<i>HNRNPUL1</i>	C161A	P54Q	Familial	Yes	11.8	No
<i>EPB41L3</i>	G1295A/G968A	R432H/R323H	Sporadic	No	29.2	No
<i>EZR</i>	C1714T	R572W	Sporadic	Yes	26.7	No
<i>GRSF1</i>	A364G	K122E	Familial	No	14.7	Yes
<i>HNRNPA3</i>	C35G	P12R	Familial	Yes	16.9	No
<i>HNRNPU</i>	C1202T/C1259T	S401L/S420L	Familial	Yes	35	Yes
<i>HSPA5</i>	G76A	D26N	Sporadic	No	18.6	No
<i>ILF3</i>	C1445T	S482L	Familial	Yes	12.4	No
<i>PA2G4</i>	A544G	I182V	Familial	Yes	11.2	No
<i>SRPK2</i>	G889C/G922C	A297P/A308P	Familial	Yes	20.4	No
<i>XRCC6</i>	T893C/T1043C/T920C	M298T/M348T/M307T	Sporadic	Yes	20	No
<i>XRCC6</i>	G1615A/G1765A/G1642A	G539R/G589R/G548R	Sporadic	Yes	11.2	No

are dependent on protein-protein interaction via PrLDs in combination with protein-RNA interaction via RRM (Harrison and Shorter, 2017). It is proposed that mutations in PrLDs or RRM can affect this interaction and may increase the probability of transition to pathological aggregation. In support of this, a significant number of mutations already associated with ALS, which occur in RRM-containing proteins with PrLDs, cluster in or close to the PrLD or the RRM and make the protein more aggregation prone (Harrison and Shorter, 2017). A prediction of this model is that mutations in multiple proteins may act in synergy to produce aggregation. Consistent with this 76% of the variants we identified in RRM-containing proteins with PrLDs are within a low complexity sequence or a RRM.

We found that rare deleterious variants in G4C2-repeat-RNA binding partners act synergistically with *C9ORF72* expansions to shorten disease duration. This is consistent with work from our group and others providing evidence for sequestration of these proteins by repeat-RNA in *C9ORF72*-ALS cases (Cooper-Knock et al., 2014, 2015a). Moreover, we identified rare deleterious variants in these proteins in patients without *C9ORF72* expansions suggesting that dysfunction of G4C2 binding partners could be pathogenic in the absence of *C9ORF72* expansions. Other mechanisms of *C9ORF72*-ALS pathogenesis have been highlighted in the literature, but our findings support the relative importance of the repeat-RNA sequestration hypothesis. We have shown that, based on proposed RNA toxicity, we could select candidate genes and identify novel ALS genetic variants.

It is noteworthy that if all identified mutations are considered together then there is no correlation between variant-load and clinical phenotype. This probably reflects the diversity of mechanisms affected. To understand oligogenic inheritance, our data suggest that mutations will have to be understood as

acting synergistically only within groups of functionally related genes/proteins.

Many of the variants identified potentially represent novel causative ALS genes, but we were not able to demonstrate segregation in families due to an absence of available samples. In certain cases the clustering of mutations with changes identified in Project MinE and the ALS Variant Server is highly suggestive of true pathogenicity. Most compelling are examples where we have identified more than one patient with a candidate mutation. Mutations that we believe are most likely to represent novel ALS variants and genes will now be discussed.

SLC1A3

SLC1A3 encodes excitatory amino acid transporter 1 (EAAT1) which is a glial glutamate transporter and also a G4C2 binding partner. Mutations of *SLC1A3* are a cause of episodic ataxia type 6 (EA6). The proposed mechanism is excitotoxicity via loss of glutamate uptake—excitotoxicity has also been proposed as a pathophysiological mechanisms in ALS (Cooper-Knock et al., 2013). Of the mutations associated with EA6, a p.C186S mutation in transmembrane segment 4 is the closest to both of our identified variants: p.N216S and p.F170L (Table 4). Transmembrane segment 4 has been associated with inter-subunit contact to stabilize the trimeric structure of the transporter (Yernool et al., 2004). The p.N216S mutations occur in a eukaryotic specific insertion between transmembrane domains 4b and 4c. The p.F170L mutation occurs in transmembrane domain 4A. Interestingly, while complete loss of *SLC1A3* function leads to a severe phenotype with progressive ataxia (Jen et al., 2005), mutation in transmembrane segment 4 has been associated with partial loss of function and variable penetrance (de Vries et al., 2009) which is consistent with a late onset disease such

as ALS. It is noteworthy that Project MinE identified an additional ALS patient with a rare (ExAC frequency <1/10,000 control alleles) mutation within the 4A transmembrane region.

EEF1G

EEF1G encodes a component of the elongation factor-1 (EF1) complex involved in the elongation phase of protein translation which is a G4C2 binding partner. The EEF1G subunit is not proposed to have a direct role in translation (Fan et al., 2010), but co-immunoprecipitates with tubulin (Janssen and Moller, 1988) and has been observed to bind mRNA directly (Al-Maghrebi et al., 2002). This is consistent with a role for EEF1G in anchoring and translation of mRNAs in cytoskeleton bound ribosomes (Corbi et al., 2010). Translation at sites distant from the nucleus is particularly relevant in neurons and in large motor neurons in particular. We have identified two patients with mutations in the C-terminal domain of EEF1G; p.D403E and p.R327C (Table 4). Project MinE identified an additional ALS patient with a T902C variant in exon 8, the same exon as the C979T change we have identified.

XRCC6

XRCC6 is a component of the non-homologous end joining (NHEJ) complex involved in repair of double stranded DNA breaks and is a G4C2 binding partner. Two patients were identified with rare deleterious variants in *XRCC6*: p.M348T and p.G589R (Table 4). Both variants occur within DNA binding domains, therefore both variants could conceivably lead to loss of function which is consistent with our disease model. Deletion of *XRCC6* in mice leads to premature aging without an increased rate of neoplasm (Li et al., 2007). This is consistent with observations in ALS and indeed impairment of NHEJ has been previously implicated in ALS (Sama et al., 2014).

PPARGC1B

PPARGC1B is a transcription factor with roles in energy metabolism and mitochondrial biogenesis and a RRM-containing protein with a PrLD. We identified three young sporadic patients with rare deleterious variants in *PPARGC1B*: p.P385H, p.S589G, and p.E434K (Table 3). Two of the variants identified lie within exon 4 either within or close to a low complexity region containing glutamic acid repeats. The p.E434K variant is actually within the glutamic acid repeats region and the same genetic change is observed in an additional three familial ALS cases within the ALS Variant Server. It seems likely that the variants we have identified and those found in the ALS Variant Server affect the function of the PrLD within *PPARGC1B*, leading to an increased risk of pathological aggregation.

C9ORF72

A rare predicted deleterious variant was identified in *C9ORF72* in a patient who also carries a G4C2-repeat expansion. From a single patient it is not possible to determine whether there was synergy between the variant and the expansion but it is noteworthy that the patient identified suffered rapidly

progressive disease: death occurred in 12 months from first symptom onset. In our population this is within the 10% most rapidly progressive *C9ORF72*-ALS patients (Cooper-Knock et al., 2012). If this variant is pathogenic and synergistic with the G4C2-repeat expansion, then it provides some insight into the pathogenesis of *C9ORF72*-ALS. A variant in *C9ORF72* could not recapitulate the proposed gain-of-function toxicity attributed to the G4C2-repeat, but it could potentially cause loss-of-function highlighting the relative importance of proposed haploinsufficiency due to G4C2-repeat expansion.

CONCLUSION

For the first time we have provided evidence for an oligogenic model of ALS in which rare variants act synergistically within discrete pathways. We have highlighted RRM-containing proteins with PrLDs and illustrated how mutations in G4C2-binding partners might exacerbate sequestration of the same proteins by repeat-RNA transcribed from the *C9ORF72* expansion. Several of the mutations we identified are candidate novel ALS genes and we have highlighted the examples of *SLC1A3*, *EEF1G*, *XRCC6*, and *PPARGC1B*. Our findings have significant implications for the design of ALS disease models and therapeutics.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of South Sheffield Research Ethics Committee with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the South Sheffield Research Ethics Committee.

AUTHOR CONTRIBUTIONS

JC-K, AH, GH, JK, and PS were responsible for the conception and design of the study. JC-K, PH, MW, TW, MK, CM, PI, and PS were responsible for data acquisition. JC-K, HR, and IN were responsible for analysis of data. JC-K, AH, GH, JK, and PS were responsible for interpretation of data. The Project MinE ALS Sequencing consortium was involved in data acquisition and analysis. All authors were responsible for revising the manuscript and approving the final version for publication. All authors are responsible for the accuracy and integrity of the work. All authors, including members of the Project MinE ALS Sequencing consortium, meet the four ICMJE authorship criteria.

PROJECT MINE ALS SEQUENCING CONSORTIUM

Ahmad Al Kheifat, Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; **Ammar Al-Chalabi**, Maurice Wohl Clinical Neuroscience Institute,

King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; **Nazli Basak**, Suna and Inan Kirac Foundation, Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey; **Ian Blair**, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW, Australia; **Annelot Dekker**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Orla Hardiman**, Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Republic of Ireland; Department of Neurology, Beaumont Hospital, Dublin, Republic of Ireland; **Winston Hide**, Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, United Kingdom; Biostatistics Department, Harvard School of Public Health, Boston, MA, United States; **Alfredo Iacoangeli**, Department of Biostatistics, IoPPN, King's College London, London, United Kingdom; **Kevin Kenna**, Department of Neurology, University of Massachusetts Medical School, Worcester, MA, United States; **John Landers**, Department of Neurology, University of Massachusetts Medical School, Worcester, MA, United States; **Russel McLaughlin**, Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland; **Jonathan Mill**, Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; University of Exeter Medical School, Exeter University, St. Luke's Campus, Magdalen Street, Exeter EX1 2LU, United Kingdom; **Bas Middelkoop**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Mattieu Moisse**, KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium; VIB, Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium; University Hospitals Leuven, Department of Neurology, Leuven, Belgium; **Jesus Mora Pardina**, Hospital San Rafael, Madrid, Spain; **Karen Morrison**, Faculty of Medicine, University of Southampton, Southampton, United Kingdom; **Stephen Newhouse**, Department of Biostatistics, IoPPN, King's College London, London, United Kingdom; Biomedical Research Centre for Mental Health, IoPPN, King's College London, London, United Kingdom; **Sara Pulit**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Aleksey Shatunov**, Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; **Chris Shaw**, Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; **William Sproviero**, Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, United Kingdom; **Gijs Tazelaar**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Philip van Damme**, KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND),

B-3000 Leuven, Belgium, VIB, Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium; University Hospitals Leuven, Department of Neurology, Leuven, Belgium; **Leonard van den Berg**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Rick van der Spek**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Kristelvan Eijk**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Michael van Es**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Wouter van Rheenen**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Joke van Vugt**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Jan Veldink**, Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, Netherlands; **Maarten Kooyman**, SURFsara, Amsterdam, Netherlands; **Jonathan Glass**, Emory University, Atlanta, United States; **Wim Robberecht**, KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium; VIB, Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium; University Hospitals Leuven, Department of Neurology, Leuven, Belgium; **Marc Gotkine**, Hadassah University Hospital, Jerusalem, Israel; **Vivian Drory**, Tel-Aviv Medical Center, Tel-Aviv, Israel; **Matthew Kiernan**, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW, Australia; **Miguel Mitne Neto**, Universidade de São Paulo, Brazil; **Mayana Ztaz**, Universidade de São Paulo, Brazil; **Philippe Couratier**, Université de Limoges, France; **Philippe Corcia**, Université François-Rabelais, Tours, France; **Vincenzo Silani**, IRCCS Istituto Auxologico Italiano, Milan, Italy; **Adriano Chio**, Università degli Studi di Torino, Turin, Italy; **Mamede de Carvalho**, Instituto de Medicina Molecular, University of Lisbon, Lisbon, Portugal; **Susana Pinto**, Instituto de Medicina Molecular, University of Lisbon, Lisbon, Portugal; **Alberto Garcia Redondo**, Hospital Carlos III, Madrid, Spain; **Peter Andersen**, Umeå University, Umeå, Sweden; **Markus Weber**, Kantonsspital St. Gallen, St. Gallen, Switzerland; **Nicola Ticozzi**, IRCCS Istituto Auxologico Italiano, Milan, Italy.

ACKNOWLEDGMENTS

The authors would like to thank the ALS Variant Server (als.umassmed.edu) which is supported by funds from NIH/NINDS (1R01NS065847), AriSLA (EXOMEFALS, NOVALS), the ALS Association, and the MND Association. We acknowledge grants from EU Framework 7 (Euro-Motor), and the JPND/MRC SOPHIA, STRENGTH and ALS-CarE projects. JC-K holds a NIHR Clinical Lectureship and PS is supported as an NIHR Senior Investigator. This work was also supported by the NIHR Sheffield Biomedical Research Centre and the

Sheffield NIHR Clinical Research Facility. Biosample collection was supported by the MND Association and the Wellcome Trust (PS). We are very grateful to those ALS patients and control subjects who generously donated biosamples to contribute to this research.

REFERENCES

- Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. *Cell* 137, 146–158. doi: 10.1016/j.cell.2009.02.044
- Al-Chalabi, A., Calvo, A., Chio, A., Colville, S., Ellis, C. M., Hardiman, O., et al. (2014). Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. *Lancet Neurol.* 13, 1108–1113. doi: 10.1016/S1474-4422(14)70219-4
- Al-Chalabi, A., Hansen, V. K., Simpson, C. L., Xi, J., Hosler, B. A., Powell, J. F., et al. (2003). Variants in the ALS2 gene are not associated with sporadic amyotrophic lateral sclerosis. *Neurogenetics* 4, 221–222. doi: 10.1007/s10048-003-0152-1
- Al-Maghrebi, M., Brulé, H., Padkina, M., Allen, C., Holmes, W. M., and Zehner, Z. E. (2002). The 3' untranslated region of human vimentin mRNA interacts with protein complexes containing eEF-1gamma and HAX-1. *Nucleic Acids Res.* 30, 5017–5028. doi: 10.1093/nar/gkf656
- Choi, J., Dixon, D. M., Dansithong, W., Abdallah, W. F., Roos, K. P., Jordan, M. C., et al. (2016). Muscleblind-like 3 deficit results in a spectrum of age-associated pathologies observed in myotonic dystrophy. *Sci. Rep.* 6:30999. doi: 10.1038/srep30999
- Cooper-Knock, J., Hewitt, C., Highley, J. R., Brockington, A., Milano, A., Man, S., et al. (2012). Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. *Brain* 135, 751–764. doi: 10.1093/brain/awr365
- Cooper-Knock, J., Higginbottom, A., Stopford, M. J., Highley, J. R., Ince, P. G., Wharton, S. B., et al. (2015a). Antisense RNA foci in the motor neurons of C9ORF72-ALS patients are associated with TDP-43 proteinopathy. *Acta Neuropathol.* 130, 63–75. doi: 10.1007/s00401-015-1429-9
- Cooper-Knock, J., Jenkins, T., and Shaw, P. J. (2013). Clinical and molecular aspects of motor neuron disease. *Colloquium Ser. Genomic Mol. Med.* 2, 1–60. doi: 10.4199/C00093ED1V01Y201309GMM004
- Cooper-Knock, J., Kirby, J., Highley, R., and Shaw, P. J. (2015b). The spectrum of C9ORF72-mediated neurodegeneration and amyotrophic lateral sclerosis. *Neurotherapeutics* 12, 326–339. doi: 10.1007/s13311-015-0342-1
- Cooper-Knock, J., Walsh, M. J., Higginbottom, A., Highley, J. R., Dickman, M. J., Edbauer, D., et al. (2014). Sequestration of multiple RNA Recognition Motif-containing proteins by C9ORF72 repeat expansions. *Brain* 137, 2040–2051. doi: 10.1093/brain/awu120
- Corbi, N., Batassa, E. M., Pisani, C., Onori, A., Di Certo, M. G., Strimpakos, G., et al. (2010). The eEF1gamma subunit contacts RNA polymerase II and binds vimentin promoter region. *PLoS ONE* 5:e14481. doi: 10.1371/journal.pone.0014481
- Couthouis, J., Hart, M. P., Shorter, J., DeJesus-Hernandez, M., Erion, R., Oristano, R., et al. (2011). A yeast functional screen predicts new candidate ALS disease genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20881–20890. doi: 10.1073/pnas.1109434108
- DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., et al. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 72, 245–256. doi: 10.1016/j.neuron.2011.09.011
- de Vries, B., Mamsa, H., Stam, A. H., Wan, J., Bakker, S. L., Vanmolkot, K. R., et al. (2009). Episodic ataxia associated with EAAT1 mutation C186S affecting glutamate reuptake. *Arch. Neurol.* 66, 97–101. doi: 10.1001/archneurol.2008.535
- Dixon, D. M., Choi, J., El-Ghazali, A., Park, S. Y., Roos, K. P., Jordan, M. C., et al. (2015). Loss of muscleblind-like 1 results in cardiac pathology and persistence of embryonic splice isoforms. *Sci. Rep.* 5:9042. doi: 10.1038/srep09042
- Fan, Y., Schlierf, M., Gaspar, A. C., Dreux, C., Kpebe, A., Chaney, L., et al. (2010). Drosophila translational elongation factor-1gamma is modified in response to DOA kinase activity and is essential for cellular viability. *Genetics* 184, 141–154. doi: 10.1534/genetics.109.109553
- Harrison, A. F., and Shorter, J. (2017). RNA-binding proteins with prion-like domains in health and disease. *Biochem. J.* 474, 1417–1438. doi: 10.1042/BCJ20160499
- Hirano, M., Quinzii, C. M., Mitsumoto, H., Hays, A. P., Roberts, J. K., Richard, P., et al. (2011). Senataxin mutations and amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler.* 12, 223–227. doi: 10.3109/17482968.2010.545952
- Janssen, G. M., and Moller, W. (1988). Elongation factor 1 beta gamma from *Artemia*. Purification and properties of its subunits. *Eur. J. Biochem.* 171, 119–129. doi: 10.1111/j.1432-1033.1988.tb13766.x
- Jen, J. C., Wan, J., Palos, T. P., Howard, B. D., and Baloh, R. W. (2005). Mutation in the glutamate transporter EAAT1 causes episodic ataxia, hemiplegia, and seizures. *Neurology* 65, 529–534. doi: 10.1212/01.WNL.0000172638.58172.5a
- Johnson, J. O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V. M., Trojanowski, J. Q., et al. (2010). Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* 68, 857–864. doi: 10.1016/j.neuron.2010.11.036
- Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315. doi: 10.1038/ng.2892
- Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T., et al. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536, 285–291. doi: 10.1038/nature19057
- Li, H., Vogel, H., Holcomb, V. B., Gu, Y., and Hasty, P. (2007). Deletion of Ku70, Ku80, or both causes early aging without substantially increased cancer. *Mol. Cell. Biol.* 27, 8205–8214. doi: 10.1128/MCB.00785-07
- Luigetti, M., Lattante, S., Conte, A., Romano, A., Zollino, M., Marangi, G., et al. (2013). A novel compound heterozygous ALS2 mutation in two Italian siblings with juvenile amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Frontotemp. Degener.* 14, 470–472. doi: 10.3109/21678421.2012.756036
- March, Z. M., King, O. D., and Shorter, J. (2016). Prion-like domains as epigenetic regulators, scaffolds for subcellular organization, and drivers of neurodegenerative disease. *Brain Res.* 1647, 9–18. doi: 10.1016/j.brainres.2016.02.037
- Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., et al. (2015). Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* 163, 123–133. doi: 10.1016/j.cell.2015.09.015
- Münch, C., Sedlmeier, R., Meyer, T., Homberg, V., Sperfeld, A. D., Kurt, A., et al. (2004). Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. *Neurology* 63, 724–726. doi: 10.1212/01.WNL.0000134608.83927.B1
- Orrell, R. W., Habgood, J. J., Malaspina, A., Mitchell, J., Greenwood, J., Lane, R. J., et al. (1999). Clinical characteristics of SOD1 gene mutations in UK families with ALS. *J. Neurol. Sci.* 169, 56–60. doi: 10.1016/S0022-510X(99)00216-6
- Ravits, J. (2014). Focality, stochasticity and neuroanatomic propagation in ALS pathogenesis. *Exp. Neurol.* 262(Pt B), 121–126. doi: 10.1016/j.expneurol.2014.07.021
- Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 72, 257–268. doi: 10.1016/j.neuron.2011.09.010
- Salgado, D., Bellgard, M. I., Desvignes, J. P., and Bérout, C. (2016). How to identify pathogenic mutations among all those variations: variant annotation and filtration in the genome sequencing era. *Hum. Mutat.* 37, 1272–1282. doi: 10.1002/humu.23110
- Sama, R. R., Ward, C. L., and Bosco, D. A. (2014). Functions of FUS/TLS from DNA repair to stress response: implications for ALS. *ASN Neuro.* 6:1759091414544472. doi: 10.1177/1759091414544472

SUPPLEMENTARY MATERIAL

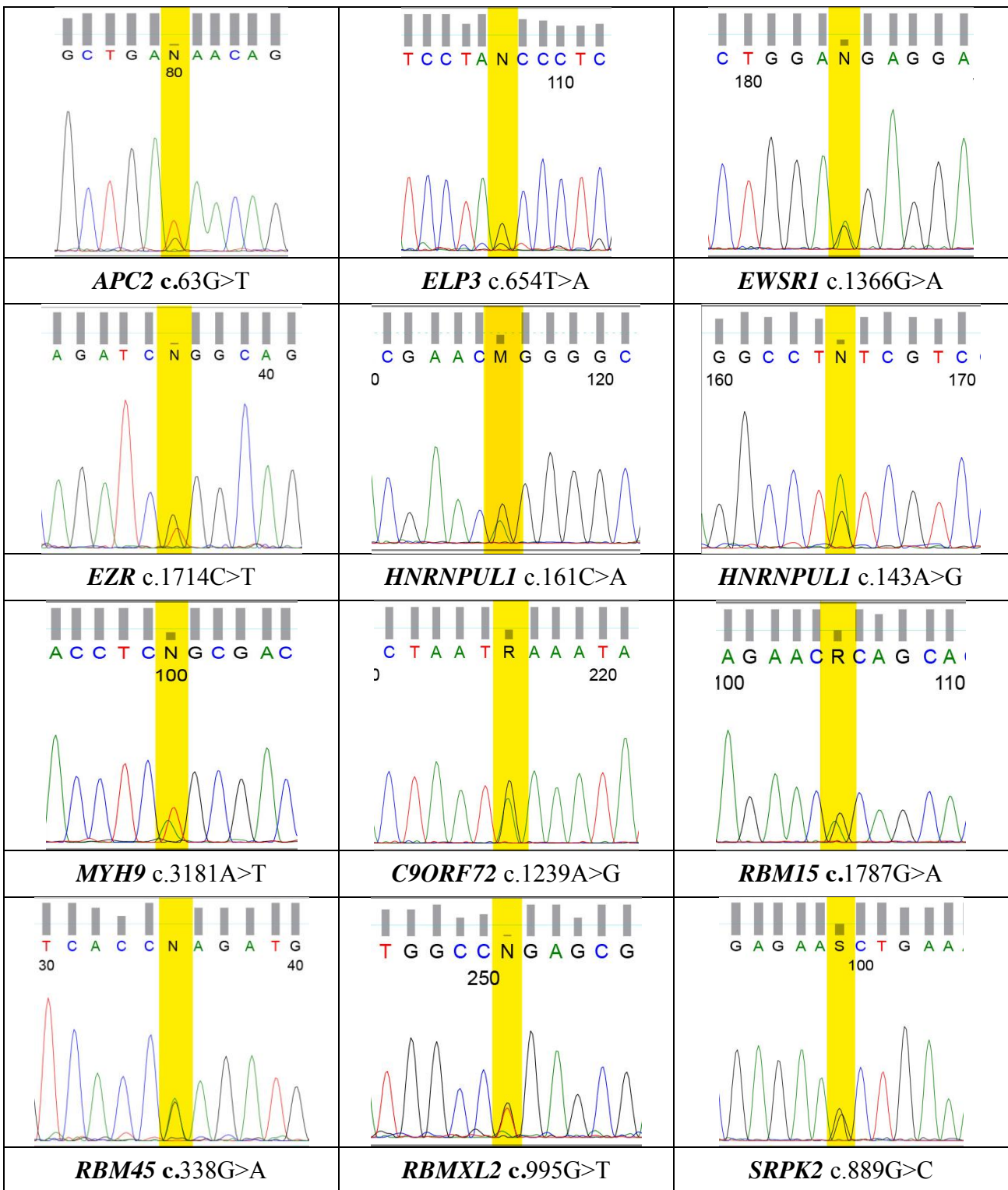
The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnmol.2017.00370/full#supplementary-material>

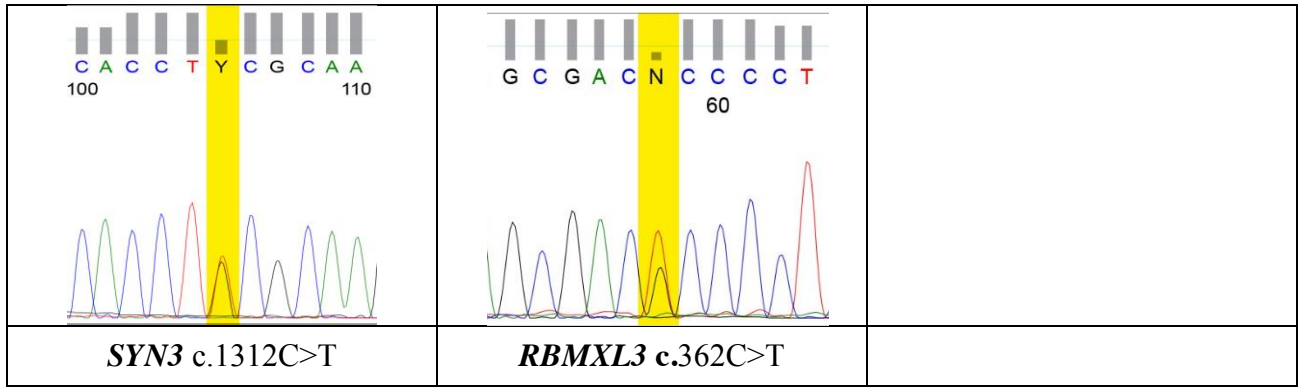
- Simpson, C. L., Lemmens, R., Miskiewicz, K., Broom, W. J., Hansen, V. K., van Vught, P. W., et al. (2009). Variants of the elongator protein 3 (ELP3) gene are associated with motor neuron degeneration. *Hum. Mol. Genet.* 18, 472–481. doi: 10.1093/hmg/ddn375
- van Blitterswijk, M., van Es, M. A., Hennekam, E. A., Dooijes, D., van Rheenen, W., Medic, J., et al. (2012). Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 21, 3776–3784. doi: 10.1093/hmg/dds199
- van Es, M. A., Veldink, J. H., Saris, C. G., Blauw, H. M., van Vught, P. W., Birve, A., et al. (2009). Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* 41, 1083–1087. doi: 10.1038/ng.442
- van Rheenen, W., Shatunov, A., Dekker, A. M., McLaughlin, R. L., Diekstra, F. P., Pulit, S. L., et al. (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat. Genet.* 48, 1043–1048. doi: 10.1038/ng.3622
- Yernool, D., Boudker, O., Jin, Y., and Gouaux, E. (2004). Structure of a glutamate transporter homologue from *Pyrococcus horikoshii*. *Nature* 431, 811–818. doi: 10.1038/nature03018

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Cooper-Knock, Robins, Niedermoser, Wyles, Heath, Higginbottom, Walsh, Kazoka, Project MinE ALS Sequencing Consortium, Ince, Hautbergue, McDermott, Kirby and Shaw. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplementary Figure 1. Variants validated by Sanger Sequencing.





Supplementary Table 1: Complete list of sequenced genes

Gene	Reason for inclusion in screen
SOD1	Known ALS Gene
FUS	Known ALS Gene
TARDBP	Known ALS Gene
C9ORF72	Known ALS Gene
OPTN	Known ALS Gene
CHMP2B	Known ALS Gene
UBQLN2	Known ALS Gene
UBQLN1	Known ALS Gene
SETX	Known ALS Gene
VAPB	Known ALS Gene
VCP	Known ALS Gene
ATXN2	Known ALS Gene
ANG	Known ALS Gene
SQSTM1	Known ALS Gene
FIG4	Known ALS Gene
DCTN1	Known ALS Gene
KIFAP3	Known ALS Gene
EWSR1	Known ALS Gene
CNTF	Known ALS Gene
ELP3	Known ALS Gene
TAF15	Known ALS Gene
EPHA4	Known ALS Gene
PFN1	Known ALS Gene
ATXN2	Known ALS Gene
NEFH	Known ALS Gene
MATR3	Known ALS Gene
SEPT2	G4C2 binding partner
APEX1	G4C2 binding partner
ATP5B	G4C2 binding partner
CALD1	G4C2 binding partner
CALM1	G4C2 binding partner
CIRBP	G4C2 binding partner
CPSF6	G4C2 binding partner
CSDA	G4C2 binding partner
DNAJC9	G4C2 binding partner
EEF1A1	G4C2 binding partner
EEF1A2	G4C2 binding partner
EEF1B2	G4C2 binding partner
EEF1D	G4C2 binding partner
EEF1G	G4C2 binding partner
EIF4H	G4C2 binding partner
EPB41L3	G4C2 binding partner
EZR	G4C2 binding partner
GLUD1	G4C2 binding partner
GRSF1	G4C2 binding partner
HDGF	G4C2 binding partner
HDGFRP3	G4C2 binding partner
HMGA1	G4C2 binding partner
HMGB1	G4C2 binding partner
HNRNPA0	G4C2 binding partner
HNRNPA1	G4C2 binding partner
HNRNPA2B1	G4C2 binding partner
HNRNPA3	G4C2 binding partner
HNRNPAB	G4C2 binding partner

HNRNPC	G4C2 binding partner
HNRNPD	G4C2 binding partner
HNRNPF	G4C2 binding partner
HNRNPH1	G4C2 binding partner
HNRNPH2	G4C2 binding partner
HNRNPH3	G4C2 binding partner
HNRNPR	G4C2 binding partner
HNRNPU	G4C2 binding partner
HNRNPUL1	G4C2 binding partner
HNRPDL	G4C2 binding partner
HSPA1L	G4C2 binding partner
HSPA5	G4C2 binding partner
HSPA8	G4C2 binding partner
ILF2	G4C2 binding partner
ILF3	G4C2 binding partner
KCTD12	G4C2 binding partner
MARCKS	G4C2 binding partner
MSN	G4C2 binding partner
MYH9	G4C2 binding partner
MYL6	G4C2 binding partner
NACA	G4C2 binding partner
NARS	G4C2 binding partner
NCL	G4C2 binding partner
NPM1	G4C2 binding partner
NUDT21	G4C2 binding partner
PA2G4	G4C2 binding partner
PABPC1	G4C2 binding partner
PPIB	G4C2 binding partner
PSIP1	G4C2 binding partner
PURA	G4C2 binding partner
PURB	G4C2 binding partner
RBM3	G4C2 binding partner
RBMX	G4C2 binding partner
RPL22	G4C2 binding partner
RPL26	G4C2 binding partner
RPLP2	G4C2 binding partner
RPS12	G4C2 binding partner
SARS	G4C2 binding partner
SERBP1	G4C2 binding partner
SFPQ	G4C2 binding partner
SFRS1	G4C2 binding partner
SFRS2	G4C2 binding partner
SFRS3	G4C2 binding partner
SFRS5	G4C2 binding partner
SFRS6	G4C2 binding partner
SFRS7	G4C2 binding partner
SFRS9	G4C2 binding partner
SLC1A3	G4C2 binding partner
SRP14	G4C2 binding partner
SRP9	G4C2 binding partner
SRPK2	G4C2 binding partner
SSB	G4C2 binding partner
SUB1	G4C2 binding partner
SYNCRIP	G4C2 binding partner
SYT1	G4C2 binding partner
THOC4	G4C2 binding partner
TPM1	G4C2 binding partner

TPM2	G4C2 binding partner
TPM3	G4C2 binding partner
TPM4	G4C2 binding partner
TRMT112	G4C2 binding partner
TUBA1B	G4C2 binding partner
TUBB	G4C2 binding partner
U2AF2	G4C2 binding partner
XRCC5	G4C2 binding partner
XRCC6	G4C2 binding partner
YBX1	G4C2 binding partner
YWHAE	G4C2 binding partner
YWHAG	G4C2 binding partner
YWHAH	G4C2 binding partner
AC015631.1	RNA binding protein with prion-like domain
AC021224.2	RNA binding protein with prion-like domain
AC021534.1	RNA binding protein with prion-like domain
AC021593.2	RNA binding protein with prion-like domain
AC027139.2	RNA binding protein with prion-like domain
AC132219.2	RNA binding protein with prion-like domain
CELF1	RNA binding protein with prion-like domain
CELF4	RNA binding protein with prion-like domain
CNOT4	RNA binding protein with prion-like domain
CPEB2	RNA binding protein with prion-like domain
CSTF2	RNA binding protein with prion-like domain
DAZL	RNA binding protein with prion-like domain
EIF3B	RNA binding protein with prion-like domain
G3BP1	RNA binding protein with prion-like domain
G3BP2	RNA binding protein with prion-like domain
HNRNPA1L2	RNA binding protein with prion-like domain
HNRNPCL1	RNA binding protein with prion-like domain
HNRNPL	RNA binding protein with prion-like domain
HNRNPM	RNA binding protein with prion-like domain
HNRPLL	RNA binding protein with prion-like domain
HTATSF1	RNA binding protein with prion-like domain
IGF2BP1	RNA binding protein with prion-like domain
IGF2BP2	RNA binding protein with prion-like domain
IGF2BP3	RNA binding protein with prion-like domain
LARP7	RNA binding protein with prion-like domain
MKI67IP	RNA binding protein with prion-like domain
MSI1	RNA binding protein with prion-like domain
MSI2	RNA binding protein with prion-like domain
MTHFSD	RNA binding protein with prion-like domain
MYEF2	RNA binding protein with prion-like domain
NCBP2	RNA binding protein with prion-like domain
NCBP2L	RNA binding protein with prion-like domain
NOL8	RNA binding protein with prion-like domain
NONO	RNA binding protein with prion-like domain
PABPC1L	RNA binding protein with prion-like domain
PABPC1L2A	RNA binding protein with prion-like domain
PABPC1L2B	RNA binding protein with prion-like domain
PABPC3	RNA binding protein with prion-like domain
PABPC4L	RNA binding protein with prion-like domain
PABPC4	RNA binding protein with prion-like domain
PABPC5	RNA binding protein with prion-like domain
PABPN1	RNA binding protein with prion-like domain
PABPN1L	RNA binding protein with prion-like domain
POLDIP3	RNA binding protein with prion-like domain

RBM1D	RNA binding protein with prion-like domain
RBM1E	RNA binding protein with prion-like domain
RBM1F	RNA binding protein with prion-like domain
RBM1J	RNA binding protein with prion-like domain
RBPMS	RNA binding protein with prion-like domain
RBPMS2	RNA binding protein with prion-like domain
RDBP	RNA binding protein with prion-like domain
RNPC3	RNA binding protein with prion-like domain
RNPS1	RNA binding protein with prion-like domain
ROD1	RNA binding protein with prion-like domain
RP11-658F2.1	RNA binding protein with prion-like domain
RRP7A	RNA binding protein with prion-like domain
SAFB	RNA binding protein with prion-like domain
SAFB2	RNA binding protein with prion-like domain
SART3	RNA binding protein with prion-like domain
SETD1A	RNA binding protein with prion-like domain
SETD1B	RNA binding protein with prion-like domain
SF3B4	RNA binding protein with prion-like domain
SFRS11	RNA binding protein with prion-like domain
SFRS12	RNA binding protein with prion-like domain
SFRS13A	RNA binding protein with prion-like domain
SFRS13B	RNA binding protein with prion-like domain
SFRS15	RNA binding protein with prion-like domain
SFRS2B	RNA binding protein with prion-like domain
SFRS4	RNA binding protein with prion-like domain
SLTM	RNA binding protein with prion-like domain
SNRNP35	RNA binding protein with prion-like domain
SNRNP70	RNA binding protein with prion-like domain
SNRPA	RNA binding protein with prion-like domain
SNRPB2	RNA binding protein with prion-like domain
SPEN	RNA binding protein with prion-like domain
SR140	RNA binding protein with prion-like domain
TDRD10	RNA binding protein with prion-like domain
TIA1	RNA binding protein with prion-like domain
TIAL1	RNA binding protein with prion-like domain
TNRC4	RNA binding protein with prion-like domain
TRA2A	RNA binding protein with prion-like domain
TRA2B	RNA binding protein with prion-like domain
TRNAU1AP	RNA binding protein with prion-like domain
TUT1	RNA binding protein with prion-like domain
U2AF1	RNA binding protein with prion-like domain
UHMK1	RNA binding protein with prion-like domain
ZCRB1	RNA binding protein with prion-like domain
ZNF638	RNA binding protein with prion-like domain
ZRSR1	RNA binding protein with prion-like domain
ZRSR2	RNA binding protein with prion-like domain

Supplementary Table 2: List of rare deleterious variants identified per patient including C9ORF72 genotype and phenotype

Newly Identified Rare Deleterious Variants				C9ORF72 Expansion	Number of mutations	Number of Mutations Including C9ORF72 Expansion	Age of Onset	Disease Duration	Familial / Sporadic
SLC1A3				No	1	1	36	52	Familial
ATP5B				No	1	1	44	107	Sporadic
C9ORF72				Yes	1	2	63	12	Sporadic
DCTN1	MYH9	NOL8	RBM4B	Yes	4	5	62	20	Sporadic
DCTN1				No	1	1	49	49	Sporadic
EEF1G				Yes	1	2	69	22	Familial
EEF1G				No	1	1	44	82	Sporadic
EIF3B				No	1	1	38	19	Familial
ELP3				No	1	1	58	4	Familial
EPB41L3	RBM41			No	2	2	23	116	Sporadic
EWSR1				No	1	1	61	68	Familial
EZR				No	1	1	28	48	Sporadic
GRSF1	RBM12	RBM15		Yes	3	4	67	14	Familial
HNRNPA3				No	1	1	53	165	Familial
HNRNPM				No	1	1	37	81	Sporadic
HNRNPU				Yes	1	2	69	27	Familial
HNRNPUL1				No	1	1	47	183	Familial
HSPA5	PPARGC1B			No	2	2	47	71	Sporadic
ILF3				No	1	1	NA	NA	Familial
MTHFSD	SOD1			No	2	2	46	18	Sporadic
SPEN				No	1	1	32	38	Sporadic
PA2G4				No	1	1	78	13	Familial
PABPC1L	RBMXL3			No	2	2	46	42	Sporadic
PPARGC1B				No	1	1	23	116	Sporadic
PPARGC1B				No	1	1	33	54	Sporadic
RAVER1				No	1	1	43	101	Sporadic
RBM12B				No	1	1	40	55	Sporadic
RBM15B				No	1	1	43	56	Sporadic
RBM45				Yes	1	2	50	40	Sporadic
RBMS2				No	1	1	67	27	Familial
RBMXL2				No	1	1	43	41	Sporadic
SETX				No	1	1	88	4	Familial
SETX				Yes	1	2	67	141	Familial
SLC1A3				Yes	1	2	56	27	Sporadic
SOD1				No	1	1	39	89	Sporadic
SRPK2				No	1	1	59	115	Familial
TRNAU1AP				No	1	1	51	35	Sporadic
UNC13A				No	1	1	43	21	Familial
VCP				No	1	1	49	15	Sporadic
XRCC6				No	1	1	47	51	Sporadic
XRCC6				No	1	1	39	53	Sporadic
ALS2				No	1	1	41	92	Sporadic
				Yes	0	1	58	7	Sporadic
				Yes	0	1	50	42	Familial
				Yes	0	1	50	28	Familial
				Yes	0	1	47	19	Familial
				Yes	0	1	63	43	Familial