Original research

Value of systematic genetic screening of patients with amyotrophic lateral sclerosis

Stephanie R Shepheard ,[1] Matthew D Parker,[1] Johnathan Cooper-Knock ,[1] Nick S Verber,[1] Lee Tuddenham,[1] Paul Heath,[1] Nick Beauchamp,[2] Elsie Place,[2] Elizabeth S A Sollars,[2] Martin R Turner ,[3] Andrea Malaspina,[4] Pietro Fratta,[5,6] Channa Hewamadduma,[7] Thomas M Jenkins ,[1] Christopher J McDermott ,[1] Dennis Wang,[8] Janine Kirby,[1] Pamela J Shaw ,[1,7] on behalf of the Project MINE Consortium

► Additional material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jnnp-2020-325014).

For numbered affiliations see end of article.

Correspondence to
Professor Pamela J Shaw, Sheffield Institute for Translational Neuroscience, The University of Sheffield, Sheffield, Sheffield, UK; pamela.shaw@sheffield.ac.uk

SRS, MDP and JC-K contributed equally.

JK and PJS are joint senior authors.

Received 1 September 2020
Revised 15 November 2020
Accepted 25 November 2020
Published Online First 14 February 2021

To cite: Shepheard SR, Parker MD, Cooper-Knock J, *et al*. *J Neurol Neurosurg Psychiatry* 2021;**92**:510–518.

ABSTRACT

Objective The clinical utility of routine genetic sequencing in amyotrophic lateral sclerosis (ALS) is uncertain. Our aim was to determine whether routine targeted sequencing of 44 ALS-relevant genes would have a significant impact on disease subclassification and clinical care.

Methods We performed targeted sequencing of a 44-gene panel in a prospective case series of 100 patients with ALS recruited consecutively from the Sheffield Motor Neuron Disorders Clinic, UK. All participants were diagnosed with ALS by a specialist Consultant Neurologist. 7/100 patients had familial ALS, but the majority were apparently sporadic cases.

Results 21% of patients with ALS carried a confirmed pathogenic or likely pathogenic mutation, of whom 93% had no family history of ALS. 15% met the inclusion criteria for a current ALS genetic-therapy trial. 5/21 patients with a pathogenic mutation had an additional variant of uncertain significance (VUS). An additional 21% of patients with ALS carried a VUS in an ALS-associated gene. Overall, 13% of patients carried more than one genetic variant (pathogenic or VUS). Patients with ALS carrying two variants developed disease at a significantly earlier age compared with patients with a single variant (median age of onset=56 vs 60 years, p=0.0074).

Conclusions Routine screening for ALS-associated pathogenic mutations in a specialised ALS referral clinic will impact clinical care in 21% of cases. An additional 21% of patients have variants in the ALS gene panel currently of unconfirmed significance after removing non-specific or predicted benign variants. Overall, variants within known ALS-linked genes are of potential clinical importance in 42% of patients.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is an adult-onset neurodegenerative disease characterised by progressive injury and cell death of upper and lower motor neurons in the motor cortex, brainstem and spinal cord. This leads to progressive failure of the neuromuscular system with death, usually from respiratory failure, within 2–5 years of symptom onset in most cases. Up to 50% of cases also show mild cognitive impairment, with approximately 5% progressing to clinically recognised fronto-temporal dementia (FTD).[1] While the majority of ALS cases are considered sporadic (sALS), 5%–10% have been shown to be familial, usually with autosomal dominant inheritance, and the genetic cause of approximately 60%–70% of familial ALS (fALS) cases has now been identified.[2] The most common genetic cause of ALS is due to expansion of a GGGGCC (G4C2) hexanucleotide repeat in the first intron of the *C9orf72* gene. This expansion has a frequency of 43% in fALS and 7% in sALS cases in our UK cohort,[3] which is comparable with worldwide figures of 39.3% for fALS and 7.0% for sALS.[4] Mutations in *SOD1*,[5 6] *TARDBP*,[5 7 8] and *FUS*[5 9] genes, the next most common genetic causes of ALS, have also been reported in both patients with fALS and those with sALS. Therefore, it is clear that apparently sporadic cases can also carry potentially pathogenic variants in known ALS genes. In a recent study which screened 17 ALS-related genes, 27.8% of apparently sporadic cases carried a potentially pathogenic or rare variant in a known ALS gene.[10] In addition, it was noted that 3.8% of patients also carried multiple variants, with these cases having a significantly earlier age of onset. Another recent report from an Australian sporadic ALS cohort found that one-third of patients carried a variant of interest and 7% carried two or more variants, which again was correlated with an earlier age of onset.[11] It has previously been reported that ALS is a six-step process, with genes, environment and time (in the form of ageing) contributing to disease development.[12] It was proposed that individuals with a genetic variant would require fewer steps than those without such variants. Using data from an ALS registry in Italy, this proved to be the case, with individuals carrying *C9orf72*, *TARDBP* or *SOD1* mutations showing a three-step, four-step and two-step process.[13]

Currently, only cases with a familial history of ALS, dementia or with a young age of disease onset tend to be routinely offered genetic screening in a clinical setting, at least in the UK.[14] However, with the advent of therapies targeting specific genetic forms of the disease associated with *SOD1* or *C9orf72* mutations (Biogen sponsored clinical trials

Table 1 Cohort demographics and phenotypes

Total no of patients	100
Age (years), mean (range)	60.4 (22–87)
Male	54%
Female	46%
Site of first muscle weakness	
Upper limb	40%
Lower limb	39%
Bulbar	19%
Respiratory	1%
Trunk	1%
Disease subtype	
ALS	83%
LMN predominant	12%
UMN predominant	5%
Family history of ALS	7%
Time from onset of first muscle weakness (months), mean (SD, range)	20.8 (20.4, 1–134)
Time from onset of first muscle weakness until death (months), mean (SD, range)*	28.5 (21.0, 7–85)
ALSFRS-R, mean (range) at first assessment	36.9 (15–47)
Revised El-Escorial criteria at first assessment	
Clinically probable laboratory supported	36%
Clinically probable	30%
Clinically suspected	15%
Clinically definite	13%
Clinically possible	6%

ALS, amyotrophic lateral sclerosis; ALSFRS-R, revised ALS-Functional Rating Score; LMN, lower motor neuron; UMN, upper motor neuron.

of BIIB067 and BIIB078), this raises the question as to whether all patients with ALS should be offered genetic screening.

Currently, only two neuroprotective therapies are available for patients with ALS, riluzole[15] and edaravone[16], and these agents have only very modest effects on the course of the disease. New approaches are required to improve outcomes in ALS clinical trials and key to this is improved disease stratification, an approach which has been very successful in several areas of oncology.[17]

'*A Multi-Centre Biomarker Research Strategy in ALS*' (AMBRoSIA) is a longitudinal biosampling programme in which newly referred patients with ALS are approached for research participation. Consenting patients have genetic screening undertaken on a research basis and samples of their blood, urine and cerebrospinal fluid are collected longitudinally, together with a skin biopsy for fibroblast culture and reprogramming. Here, we report our findings from the prospective genetic screening of a highly representative and deeply phenotyped cohort of patients with sporadic and familial ALS.

MATERIALS AND METHODS

Patient cohort and biosamples

The first 100 patients with clinically identified ALS (7 familial and 93 apparently sporadic) recruited to the Sheffield site of the AMBRoSIA programme were analysed in this study (table 1). The patients were recruited between June 2017 and October 2019. Biosamples, including whole blood for DNA extraction, were obtained following informed written consent (REC 16/LO/2136). Other than a confirmed diagnosis of ALS, no strict inclusion or exclusion criteria were adopted (see online supplemental table S5 for AMBRoSIA inclusion/exclusion criteria), so patients were not prioritised for genetic screening by clinicians. The diagnosis of ALS was made by an experienced neuromuscular neurologist (PJS, CJM, TMJ, CH) following appropriate investigations to exclude alternative diagnoses, and detailed clinical and demographic features were recorded for all patients.

Illumina targeted panel sequencing

A panel of 44 ALS, motor system and FTD-linked genes (figure 1A) was screened to diagnostic standards using targeted next-generation sequencing by the UKAS-accredited Sheffield Diagnostic Genetics Service laboratory, as part of the AMBRoSIA project. The panel was approved for familial ALS with and without FTD by the UK Genetic Testing Network steering group. DNA for the panel was captured using SureSelectXT (Design ID: 0836801) automated library preparation and libraries were sequenced on an Illumina HiSeq 2500 in rapid run mode at 2×107 bp. A mean coverage of at least 100x was obtained (online supplemental figure S1).

C9ORF72 expansion testing

Hexanucleotide repeat expansions (G4C2) in *C9ORF72* were tested by flanking PCR and, if required, fluorescent repeat primed PCR (RP-PCR) (online supplemental table S1). Fragment size analysis was performed in GeneMapper (V.3.5). Expansions were reported as normal if <30 repeats were detected, and all expansions of >30 repeats plus the classical sawtooth pattern were reported as potentially pathogenic.

ATXN2 expansion testing

ATXN2 repeat expansions (CAG) were tested by standard PCR with one fluorescently labelled primer (online supplemental table S2), followed by analysis using GeneMapper (V.3.5). ATXN2 repeats were reported as normal between 14 and 28; 35 repeats and above were reported as consistent with spinocerebellar ataxia type 2 (SCA2) and repeat lengths of 29–34, which have been associated with an increased risk of ALS, were reported as intermediate repeat lengths.[18]

Sequencing and variant analysis

Sequencing data were analysed using a clinically validated bioinformatics pipeline. Samples were checked for contamination (online supplemental figure S2). Reads were aligned with bwa[19] (V.0.7.15) to a bespoke version of the human reference (hg19). Indels were realigned with GATK[20] (V.3.7). Variants were called with GATK Haplotype Caller (V.3.7) and decomposed and normalised with vt[21] (V.0.5) and uploaded to Fabric Genomics Opal (V.6.1.8). Variants with a quality score <1500 were excluded from further analysis. Protein coding variants, only, were retained.

The pathogenicity of a variant was determined using a multifaceted approach, including manual review in Opal, population frequency and in silico software algorithms, as well as the presence or absence of variants in our clinical reporting pipeline and the ALS literature.

Fabric Genomics Opal (V.6.1.8) provides population frequency from the 100,000 Genomes Project, Exome Variant Server of NHLBI GO Exome Sequencing Project, ExAC and gnomAD. We chose to report on overall population frequency using gnomAD, the largest dataset available.[22]

In silico analysis used the Omicia Score, which combines software algorithms MutationTaster,[23] Polymorphism Phenotyping v2 (PolyPhen2),[24] Sorting Intolerant from Tolerant (SIFT)[25] and phyloP[26]—placental, primate and vertebrate. Additional

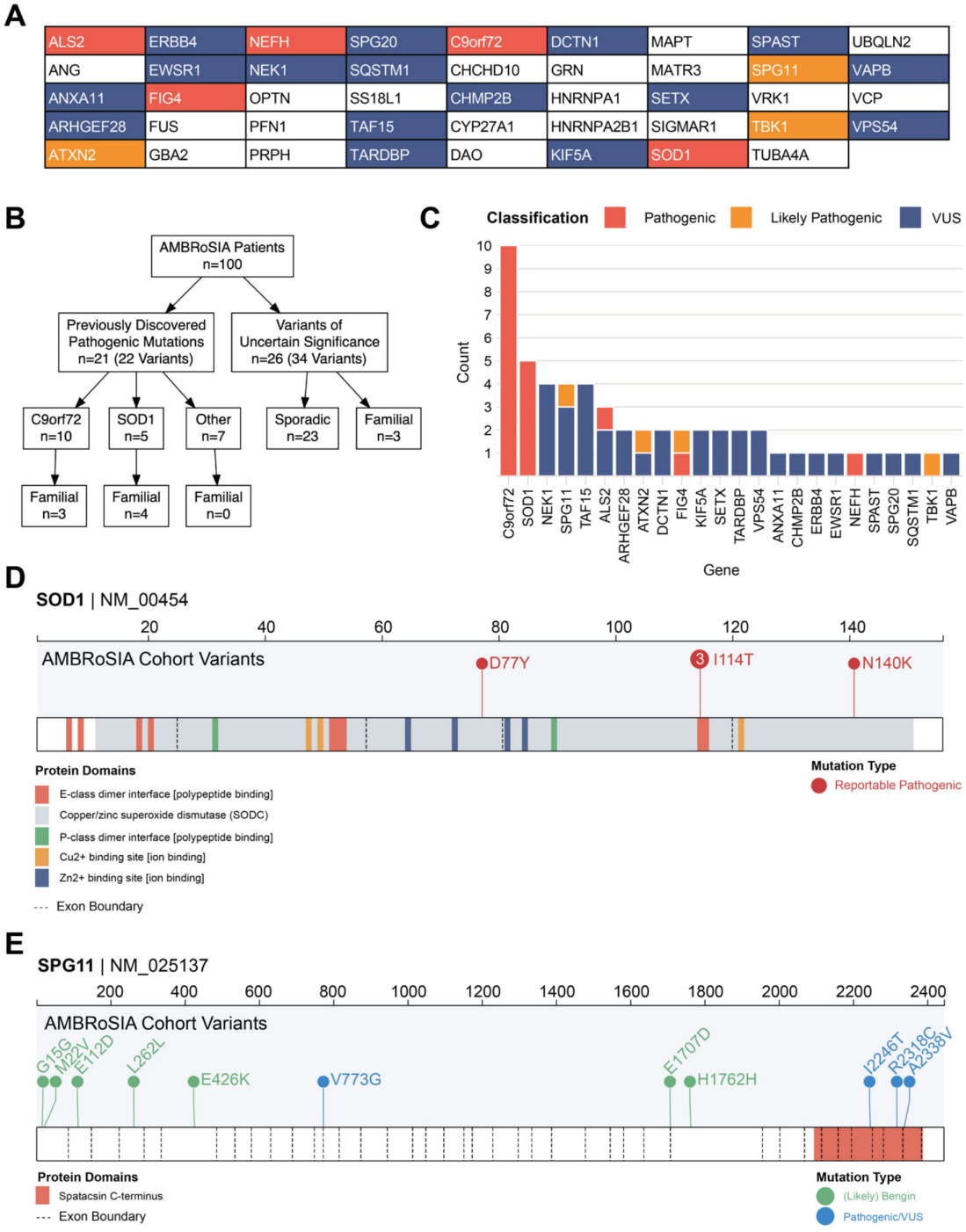

Figure 1 Clinical screening of all patients with ALS identifies pathogenic, likely pathogenic and variants of uncertain significance (VUS). (A) The 44 genes covered by the clinical ALS panel (42 profiled by next-generation sequencing and 2 by PCR). Those genes in which pathogenic (red), likely pathogenic variants (orange) or VUS (blue) were identified after filtering. No pathogenic variants or VUS were found in the genes depicted in the white boxes. Background colour of the box represents the most severe variant found in that gene. (B) Schematic of our variant analysis and filtering process. (C) Counts of reportable (red) and variants of unknown significance in ALS genes in the Sheffield AMBRoSIA cohort. (D) Five clinically reportable variants in *SOD1*. (E) 4 Variants of unknown significance were discovered in *SPG11*. Interestingly, 3 of these cluster in the C-terminal domain. Benign and likely benign mutations are conversely distributed throughout the protein. Protein domain figures created using ProteinPaint (https://pecan.stjude.cloud/proteinpaint).

algorithms used were VAAST Variant Prioritization (VVP) and Combined Annotation Dependent Depletion (CADD).[27] Classification of variant effects was based on transcripts detailed in online supplemental table S3. Potential splicing disruption was predicted using Neural Network Splice (NNSplice),[28] GeneSplicer[29] and MaxEntScan.[30]

The ClinVar database (2 January 2019[31]) was also used to determine if the variant had been previously assessed. Variants previously described as pathogenic or likely pathogenic in Opal, the genetics service clinical pipeline, or the literature (based ClinVar) were labelled as such. Variants with a lesser likelihood of pathogenicity were labelled as benign and excluded from reporting if

Table 2 Clinically established and ACMG reportable variants in ALS-associated genes: 22 changes in 21 individuals

Sample	Gene	Genomic variant	Protein change (HGVSp canonical)	Population frequency (gnomAD)	Classification
SHF-094	ALS2	chr2:g.202625839G>A	p.Thr293Ile	1.00E−05	Pathogenic
SHF-014	C9orf72	–	–	NA	Pathogenic
SHF-017	C9orf72	–	–	NA	Pathogenic
SHF-018	C9orf72	–	–	NA	Pathogenic
SHF-043	C9orf72	–	–	NA	Pathogenic
SHF-081	C9orf72	–	–	NA	Pathogenic
SHF-083	C9orf72	–	–	NA	Pathogenic
SHF-103	C9orf72	–	–	NA	Pathogenic
SHF-138	C9orf72	–	–	NA	Pathogenic
SHF-152	C9orf72	–	–	NA	Pathogenic
SHF-162	C9orf72	–	–	NA	Pathogenic
SHF-021	ATXN2	–	–	NA	Likely pathogenic
SHF-065	FIG4	chr6:g.110107620T>TA	p.Thr689AsnfsTer12	None	Likely pathogenic
SHF-080	FIG4	chr6:g.110036336T>C	p.Ile41Thr	0.00102	Pathogenic
SHF-094	NEFH	chr22:g.29885959C>T	p.Pro777Leu	3.00E−05	Pathogenic
SHF-006	SOD1	chr21:g.33039672T>C	p.Ile114Thr	5.00E−05	Pathogenic
SHF-039	SOD1	chr21:g.33039672T>C	p.Ile114Thr	5.00E−05	Pathogenic
SHF-078	SOD1	chr21:g.33039672T>C	p.Ile114Thr	5.00E−05	Pathogenic
SHF-082	SOD1	chr21:g.33040846C>A	p.Asn140Lys	0	Pathogenic
SHF-007	SOD1	chr21:g.33038821G>T	p.Asp77Tyr	None	Pathogenic
SHF-085	SPG11	chr15:g.44856883G>A	p.Ala2338Val	1.00E−05	Likely pathogenic
SHF-084	TBK1	chr12:g.64891000TGAA>T	p.Glu643del	1.00E−05	Likely pathogenic

ACMG, American College of Medical Genetics and Genomics; ALS, amyotrophic lateral sclerosis; NA, not available.

they had one of the following properties: (1) intronic or synonymous variants with no predicted effect on splice sites; (2) an allele frequency >0.001 according to gnomAD[22]; (3) previously reported as benign in ClinVar; (4) a CADD score <15.[27] Variants of uncertain significance (VUS) were determined as those that did not fit the aforementioned criteria for pathogenic/likely pathogenic or benign. CADD score is not available for indels and therefore indels were denoted as VUS if minor allele frequency (MAF) <0.001 and they passed following QC criteria: Quality Depth (QD) <2, Fisher Strand (FS) >200, ReadPosRankSum <−20 (n=43) and if the frequency in the present cohort was <0.1 (n=23). Those that are >0.001 in gnomAD (n=20) and non-coding indels were removed (n=13). Finally, we removed those that were common in controls in ProjectMine (see online supplemental table S4 for details of the ProjectMine sequencing consortium), leaving 11 indels (online supplemental figures S3 and S4).

RESULTS

Prospective genetic testing leads to identification of clinically reportable pathogenic mutations

We profiled a panel of 44 relevant genes (figure 1A) in 100 prospectively identified patients with ALS using Best Practice NHS pipelines.[32] In 21 patients (21%), we identified 22 clinically reportable pathogenic and likely pathogenic variants (table 2 and figure 1B,C). Seven of the 21 patients reported a family history of ALS in a first-degree relative, but 14 patients with a clinically reportable pathogenic mutation had apparently sporadic disease.

As expected, the most frequently identified pathogenic mutation was in *C9ORF72* in 10 patients with ALS (10% of the cohort).[3] Three of these patients had reported a family history of ALS. We identified five patients (5% of the cohort) each with a clinically reportable *SOD1* mutation (table 2 and figure 1D), four of whom had a family history of ALS. Identification of *C9ORF72* and *SOD1* mutations is particularly important because these genes are associated with ongoing genetic-therapy trials (Biogen BIIB067 and BIIB078).

We identified two patients with sporadic ALS with clinically reportable changes in *FIG4*: p.Ile41Thr and p.Thr689AsnfsTer12 (table 2). The p.Ile41Thr mutation is reported as pathogenic in ClinVar, with no conflicts. The p.Thr689AsnfsTer12 variant has been reported twice before and only in sporadic ALS, suggesting variable penetrance. A frameshift mutation is likely to lead to haploinsufficiency (online supplemental figure S5). *FIG4* variants are thought to disrupt local folding of the protein leading to a reduction in stability which inhibits the normal function of the protein in lysosomal trafficking.[33]

Four additional mutations in three patients were identified which were classified as clinically reportable (table 2). A single patient with sALS carried a p.Pro777Leu *NEFH* mutation and a p.Thr293IIe *ALS2* mutation. The p.Thr293Ile change was heterozygous, whereas *ALS2* mutations are usually considered to be autosomal recessive.[34] As a result, the clinical significance of this finding is currently uncertain. However, it is possible that the two identified changes act synergistically. A single patient harboured a p.Ala2338Val *SPG11* mutation and a further single patient carried an inframe deletion within *TBK1* p.Glu643del. *SPG11* mutations are usually considered to be recessive and therefore the clinical significance of this finding is at present uncertain. The *TBK1* change is reported in ClinVar as pathogenic. We identified one patient with an expansion of 29 CAG repeats in the *ATXN2* gene which has been reported as a risk factor for ALS.[18] The majority of cases in this cohort had 22 or 23 repeats (range 15–29).

Panel screening leads to identification of VUS

Thirty-four variants in 26 patients with ALS (3 familial and 23 sporadic) (26%) were reported in ClinVar as of uncertain significance, or absent from ClinVar and met our criteria for predicted

Table 3 Variants of unconfirmed significance with potential for pathogenicity in ALS-associated genes

Sample	Gene	Genomic variant	Protein change (HGVSp canonical)	Population frequency (gnomAD)	Deleterious score (CADD)
SHF-029	ALS2	chr2:g.202626478C>T	p.Ser80Asn	None	15.15
SHF-147	ALS2	chr2:g.202609025A>G	p.Leu709Pro	None	28.9
SHF-082	ANXA11	chr10:g.81930624G>C	p.Pro35Ala	0.00012	22.4
SHF-139	ARHGEF28	chr5:g.73181708C>T	p.Ala1030Val	7.00E−05	28.5
SHF-159	ARHGEF28	chr5:g.73200229A>T	p.Gln1387Leu	None	34
SHF-110	ATXN2	chr12:g.112036798TGCTGCTGCTGCTGCTGCTGC>T	p.Gln168AlafsTer75	2.00E−05	
SHF-129	CHMP2B	chr3:g.87289902A>G	p.Arg30Gly	0.00013	26
SHF-035	DCTN1	chr2:g.74597128C>G	p.Glu452Asp	None	25.9
SHF-127	DCTN1	chr2:g.74594023G>A	p.Arg785Trp	0.00018	32
SHF-095	ERBB4	chr2:g.212566786G>T	p.Asn465Lys	9.00E−05	17.1
SHF-159	EWSR1	chr22:g.29693883T>C	p.Met459Thr	1.00E−05	22.6
SHF-157	KIF5A	chr12:g.57970115C>T	p.Arg718Trp	None	27.3
SHF-062	KIF5A	chr12:g.57957406T>C	p.Val74Ala	None	26.2
SHF-097	NEK1	chr4:g.170398474A>C	p.Asn745Lys	0.00481	26.5
SHF-144	NEK1	chr4:g.170345826G>A	p.Gln1034Ter	None	35
SHF-138	NEK1	chr4:g.170458958A>G	c.1665+2T>C	1.00E−05	29.7
SHF-128	NEK1	chr4:g.170506573T>G	p.Asp245Ala	None	23
SHF-146	SETX	chr9:g.135205481G>A	p.Arg502Trp	0.00049	27.7
SHF-162	SETX	chr9:g.135211865C>T	p.Gly179Glu	0	26.6
SHF-009	SPAST	chr2:g.32289207T>TCGGCCC	p.Ala106_Pro107dup	3.00E−05	16.17
SHF-114	SPG11	chr15:g.44914544A>C	p.Val773Gly	0.00012	28.9
SHF-139	SPG11	chr15:g.44858099G>A	p.Arg2318Cys	3.00E−05	23.2
SHF-009	SPG11	chr15:g.44859639A>G	p.Ile2246Thr	3.00E−05	29.4
SHF-148	SPG20	chr13:g.36878530ACTTC>A	p.Glu657Ter	6.00E−05	23.5
SHF-057	SQSTM1	chr5:g.179250020G>A	p.Val90Met	4.00E−05	21.9
SHF-009	TAF15	chr17:g.34171988G>A	p.Arg562Gln	2.00E−05	18.24
SHF-029	TAF15	chr17:g.34171971CGGCTATGGAGGAGACCGAGGTGGG>C	p.Gly565_Gly572del	0.0001	20.9
SHF-092	TAF15	chr17:g.34171575GGGTGGCTATGGTGGAGACAGAAGCAGC>G	p.Tyr427_Gly435del	2.00E−05	21.6
SHF-103	TAF15	chr17:g.34149811G>A	p.Arg153Lys	1.00E−05	24.3
SHF-014	TARDBP	chr1:g.11082679A>G	p.Met405Val	0	21.9
SHF-062	TARDBP	chr1:g.11082430A>G	p.Met322Val	0	22.4
SHF-128	VAPB	chr20:g.56993275A>G	p.Thr23Ala	0	23.7
SHF-144	VPS54	chr2:g.64171737A>G	p.Val433Ala	None	18.2
SHF-141	VPS54	chr2:g.64189223G>C	p.Gln327Glu	1.00E−05	24.1

AMS, amyotrophic lateral sclerosis; CADD, Combined Annotation Dependent Depletion.

pathogenic variants (table 3). Also, 5/26 patients with a VUS had an additional pathogenic variant.

Without pedigree information or population-matched controls, it is difficult to conclusively demonstrate pathogenicity. ALS has a polygenic rare variant architecture and therefore we filtered for rare predicted pathogenic variants which we quantified as MAF <0.001 and CADD >15[27] (online supplemental figure S3). In addition, we used population-matched whole genome sequencing data from 5954 patients with ALS and 2238 controls included in Project MinE[35] to check for evidence of pathogenicity based on case:control ratio (Fisher exact test $p \leq 0.05$). VUS were excluded from further analysis if they were not significantly associated with ALS in population-matched data, unless low frequency within the whole genome sequencing dataset (<10 individuals) prevented conclusive testing.

For each identified VUS, we performed a literature search to explore the case that this was indeed a newly identified ALS-risk variant. In particular, we sought evidence that a discovered VUS was present within the same functional domain as described pathogenic variants: for *TBK1*, *NEK1* and *FIG4* this included nonsense and frameshift changes because pathogenic changes in these three genes are thought to act via loss of function.[33 36] Based on these criteria, we conclude that the following VUS are likely to be pathogenic: p.Met405Val and p.Met322Val within *TARDBP*; p.Met459Thr in *EWSR1*; p.Gln1034Ter and c.1665+2T>C in *NEK1*; p.Arg2318Cys and p.Ile2246Thr in *SPG11*. Our data are suggestive of pathogenicity, but confirmations will require further validation by future studies.

Indels have high rates of inaccurate sequencing. We removed all non-coding indels, those with low quality and those present in >10% of this cohort and present in Project MinE controls leaving two frameshift variants, seven inframe deletions and two splice region variants. *FIG4* has a frameshift in codon 689 which has not previously been reported in gnomAD (V.2.1.1—online supplemental figure S5) and *SPG20* has a frameshift close to the end of the protein at codon 1969 which has a very low population frequency (0.0000636—online supplemental figure S6). *TBK1* has an inframe deletion of a single codon, which has previously been reported in ClinVar as pathogenic (online supplemental figure S7).

A

Sample	Age at Onset	Inheritance?	Gene	Variant	Gene	Variant	Gene	Variant
SHF-009	56	S	SPAST	p.Ala106_Pro107dup	SPG11	p.Ile2246Thr	TAF15	p.Arg562Gln
SHF-094	22	S	NEFH	p.Pro777Leu	ALS2	p.Thr293Ile		
SHF-159	68	S	ARHGEF28	p.Gln1387Leu	EWSR1	p.Met459Thr		
SHF-138	66	S	C9orf72	Expansion	NEK1	c.1665+2T>C		
SHF-162	63	S	C9orf72	Expansion	SETX	p.Gly179Glu		
SHF-082	53	F	ANXA11	p.Pro35Ala	SOD1	p.Asn140Lys		
SHF-139	49	S	ARHGEF28	p.Ala1030Val	SPG11	p.Arg2318Cys		
SHF-029	62	S	ALS2	p.Ser80Asn	TAF15	p.Gly565_Gly572del		
SHF-103	57	F	C9orf72	Expansion	TAF15	p.Arg153Lys		
SHF-014	52	F	C9orf72	Expansion	TARDBP	p.Met405Val		
SHF-062	51	S	KIF5A	p.Val74Ala	TARDBP	p.Met322Val		
SHF-128	56	S	NEK1	p.Asp245Ala	VAPB	p.Thr23Ala		
SHF-144	32	S	NEK1	p.Gln1034Ter	VPS54	p.Val433Ala		
			Variant 1		Variant 2		Variant 3	

S = Sporadic F = Familial

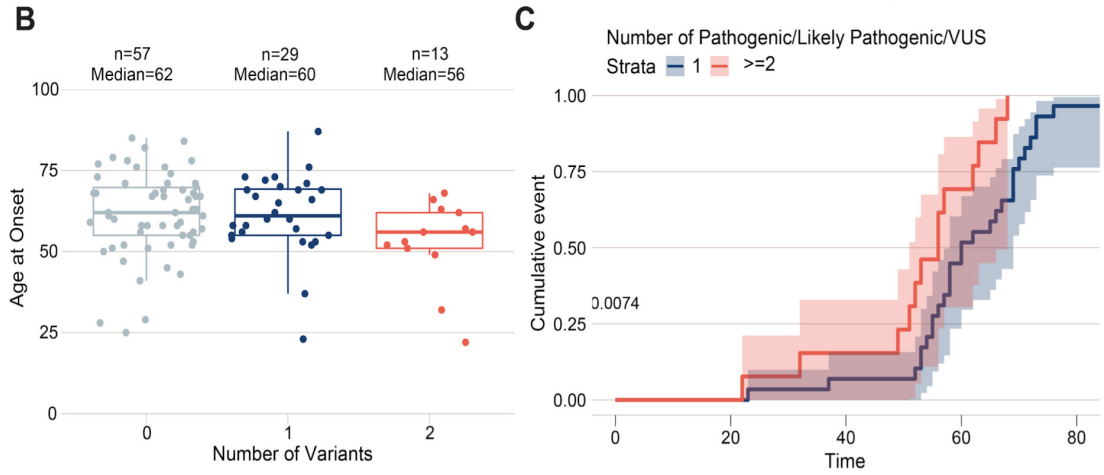

Figure 2 Increased mutation burden is associated with earlier age at onset. (A) Details of patients with two or more pathogenic, likely pathogenic or variants of uncertain significance (VUS) (n=13). (B) The number of pathogenic, likely pathogenic and VUS in each patient in the study. (C) Cumulative event plot showing those patients with two or more pathogenic, likely pathogenic or VUS have a significantly earlier age of onset.

Panel genes where no changes were found

For 20 (highlighted in figure 1A) of the 44 genes screened in our panel, no mutations (pathogenic or VUS) were identified in any member of our cohort. Many of these genetic changes are individually rare and our data are consistent with this. The absence of identified VUS in patients with sporadic ALS would be consistent with a model in which these genes harbour only high-effect, highly penetrant mutations.

Clinical correlation supports pathogenicity of VUS

In the absence of experimental validation, the biological effect of genetic variation can be inferred from correlation with clinical phenotype. The number of clinically reportable variants and VUS were combined to assign a 'mutation-load' to 42 patients with ALS (figure 2A). Patients with ALS without discovered variants were excluded because of absent information. To determine whether mutation load was clinically relevant, we compared age of onset between patients with one or two variants. Patients with ALS carrying two variants (n=13) compared with one variant developed disease at a significantly earlier age (log-rank test, p=0.0074, median age of onset=56 years vs 60 years, figure 2B,C). It is not yet possible to assess any potential effect on disease duration because 67% of this cohort are still living. Correlation with clinical phenotype is therefore consistent with a functional effect of both clinically reportable genetic variants and VUS.

DISCUSSION

An important strength of our study is its prospective nature. We performed targeted sequencing of ALS-relevant genes in 100 prospectively identified patients with ALS attending a large ALS centre in Northern England. This identified clinically reportable genetic changes in 21% of patients of whom 15 with *C9ORF72* or *SOD1* mutations would potentially be eligible for recruitment into an ongoing genetic-therapy trial. The number of clinically actionable results is likely to increase with the anticipated development of new genetic-therapy approaches for ALS. Previous genetic studies of ALS have been largely retrospective and were therefore unable to determine the utility of genetic screening in the clinic. In contrast, the present study strongly suggests that routine genetic testing should be offered in both patients with familial ALS and those with sporadic ALS, at least in our population.

We identified 22 clinically reportable mutations and 34 VUS with potential for pathogenicity in 42 patients, including 35 patients with apparently sporadic ALS. Moreover, we identified a strong correlation between mutation load and age of onset, suggesting that the majority of the VUS may be functionally important. Evaluation of the pathogenicity of individual VUS requires further study, but comparison with previously described variants in the literature suggests that we have identified novel likely pathogenic changes in *TARDBP* (two variants), *EWSR1*, *SPG11* (two changes) and *NEK1* (two variants). Genetic changes were absent from 20 genes in the panel in this cohort (figure 1A).

Mutations in RNA binding proteins are significantly associated with ALS.[37] Most prominently, *TARDBP* encodes the protein TDP-43 which is the major component of the characteristic proteinopathy in ALS. Pathogenic mutations are clustered in the C-terminal glycine-rich domain which is important for interactions with other RNA-binding proteins.[38] We have identified two patients with rare and predicted deleterious mutations in *TARDBP* which are both located in the glycine-rich domain (table 3 and online supplemental figure S8). Both changes (p.Met322Val and p.Met405Val) are absent from 125 748 population controls in gnomAD and are predicted deleterious (CADD >15), although neither has been previously reported. FET proteins including *EWSR1*, *FUS* and *TAF15* are RNA-binding proteins which have been associated with fALS. We identified several VUS within FET proteins with some evidence for pathogenicity (table 3). One patient carried a p.Met459Thr *EWSR1* mutation within the RNA binding region which contains other ALS-associated mutations.

We identified four variants within *SPG11* (figure 1E): one clinically reportable change plus two of the VUS. All four variants cluster in the spatacsin C-terminal domain, suggesting that this could represent a region of functional importance. Conversely, benign mutations in this gene tend to be more dispersed and none are found in the spatacsin C-terminal domain. We conclude that the two p.Ile2246Thr and p.Arg2318Cys VUS we identified are highly likely to be pathogenic.

Mutations in *TBK1*,[36] *NEK1*[39] and *FIG4*[33] have conclusively been shown to be loss of function (LOF) changes. In our cohort, we identified two VUS within *NEK1* which significantly disrupt the translated sequence and are therefore predicted to be highly pathogenic via a LOF mechanism (online supplemental figure S8). This included p.Gln1034Ter and c.1665+2T>C which is a splice site mutation. Both mutations are rare or absent from gnomAD and are predicted pathogenic (table 3). Similarly, we identified a clinically reportable frameshift mutation in *FIG4* (p.Thr689AsnfsTer12-FIG4) and an in-frame deletion in *TBK1* (p.Glu643del-TBK1) which has previously been reported in a single case in ClinVar (online supplemental figure S7). Our data add to the weight of evidence that this change is pathogenic and that mutations in *TBK1* are an important contributor to the genetics of sporadic ALS.[39]

VUS were identified based on universal measures of pathogenicity [27] and population frequency.[22] However, we added an additional filtering step based on case:control ratio within a population-matched whole genome sequencing dataset.[35] When combined with clinically reportable changes, the resulting VUS were associated with age of onset which suggests that these are likely to represent pathogenic changes. We confirmed previous reports that *C9ORF72* expansions are frequently associated with additional likely pathogenic mutations in other ALS genes.[40] In the present cohort, 40% of patients with pathogenic *C9ORF72* expansions carried an additional VUS. The presence of such an additional genetic variant may be one factor influencing the penetrance of *C9ORF72* mutations.

The current model of ALS is considered a multistep process in which steps constitute genetic and/or environmental exposures.[12] A consistent finding in ALS genetics is that identification of a highly penetrant genetic risk factor correlates with earlier age of onset. It follows that such mutations might be associated with fewer steps and this has been supported.[13] Our work suggests that, when correctly prioritised, VUS can also contribute steps leading to earlier age of onset and potentially function as a prognostic biomarker. Similar findings have been reported previously[10] but not in a prospective cohort; moreover, the latter study was limited to a smaller number of genes. If an index case carries one or more VUS within ALS-linked genes, then screening for VUS within family members may inform risk and even age of onset counselling. In the present cohort, an additional VUS was associated with ~4 years earlier age of onset.

In conclusion, we have performed a prospective genetic study of 100 consecutive patients with ALS attending our clinic. Our results indicate that screening of known ALS genes can lead to clinically actionable results in ~21% of patients, and in a further ~21% of patients a VUS may be discovered with potential clinical implications. As future studies expand the number of verified genetic causes of ALS, these percentages are likely to increase. We developed a pipeline for prioritising VUS using whole genome case–control cohorts to predict clinical outcomes. Although this study took place in a large tertiary referral ALS centre and requires further validation in other settings, our data suggest that all patients with ALS should, with careful counselling, be offered genetic testing, especially in light of new personalised medicine treatments in development.

Author affiliations
[1]Sheffield Institute for Translational Neuroscience, The University of Sheffield, Sheffield, UK
[2]Human Genetics, Sheffield Children's Hospital NHS Foundation Trust, Sheffield, UK
[3]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[4]Neuroscience and Trauma, Queen Mary University of London, London, UK
[5]Department of Neuromuscular Diseases, University College London Institute of Neurology, London, UK
[6]MRC Centre for Neuromuscular Diseases, University College London Institute of Neurology, London, UK
[7]Academic Directorate of Neuroscience, Department of Clinical Neurology, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK
[8]Sheffield Bioinformatics Core, The University of Sheffield, Sheffield, UK

Twitter Andrea Malaspina @NA

Acknowledgements The authors thank all study participants, their families and their carers and the SITraN and Sheffield Teaching Hospitals NHS Foundation Trust clinical staff. We also thank Timothy Freeman who contributed to the verification of variants.

Collaborators The Project MinE: Ian Blair (Centre for Motor Neuron Disease Research, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, Australia), Matthew C. Kiernan (Brain and Mind Centre, The University of Sydney, Sydney, Australia), Miguel Mitne Neto (Universidade de São Paulo, Brazil), Adriano Chio ("Rita Levi Montalcini" Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy; Azienda Ospedaliera Città della Salute e della Scienza, Torino, Italy), Ruben Cauchi (Centre for Molecular Medicine & Biobanking, University of Malta, Malta; Department of Physiology & Biochemistry, Faculty of Medicine & Surgery, University of Malta, Malta), Wim Robberecht (KU Leuven - University of Leuven, Department of Neurosciences, Leuven, Belgium; VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium), Philip van Damme (KU Leuven - University of Leuven, Department of Neurosciences, Leuven, Belgium; VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium), Philippe Corcia (Centre SLA, CHRU de Tours, Tours, France; UMR 1253, iBrain, Université de Tours, Inserm, Tours, France; Federation des Centres SLA Tours and Limoges, LITORALS, Tours, France), Philippe Couratier (Centre SLA, CHU Dupuytren Limoges, France), 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), Russell McLaughin (Complex Trait Genomics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland), Marc Gotkine (The Agnes Ginges Center for Human Neurogenetics, Hadassah-Hebrew,University Medical Center, Jerusalem, Israel), Vivian Drory (Department of Neurology Tel-Aviv Sourasky Medical Centre, Israel), Nicola Ticozzi (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy; Department of Pathophysiology and Tranplantation, 'Dino Ferrari' Center, Università degli Studi di Milano, Milano, Italy), Vincenzo Silani (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy; Department of Pathophysiology and Tranplantation, 'Dino Ferrari' Center, Università degli Studi di Milano, Milano, Italy), Jan H. Veldink (Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands), Leonard H. van den Berg (Department of Neurology, UMC Utrecht Brain Center, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands), Mamede de Carvalho (Instituto de Fisiologia, Instituto de Medicina Molecular,Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal; Department of Neurosciences, Hospital de Santa Maria-CHLN, Lisbon, Portugal), Jesus S. Mora Pardina (ALS Unit, Hospital San Rafael, Madrid, Spain), Monica Povedano (la Unitat Funcional de Motoneurona, Cap de Secció de Neurofisiologia, Servei de Neurologia, Hospital Universitario de Bellvitge-IDIBELL), Peter Andersen (Department of Clinical Science, Neurosciences, Umeå University, Sweden), Markus Weber (Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, 9007, St. Gallen, Switzerland), Nazli A. Ba_ak (Koç University, School of Medicine, KUTTAM-NDAL, Istanbul Turkey), Ammar Al-Chalabi (Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, UK), Chris Shaw (Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, UK), Pamela J. Shaw (Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK), Karen E. Morrison (School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, UK), John E. Landers (Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA), Jonathan D. Glass (Department Neurology, Emory University School of Medicine, Atlanta, GA, USA; Emory ALS Center, Emory University School of Medicine, Atlanta, GA, USA), Eran Hornsetin (Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel).

Contributors PJS and JK conceived and designed the study. JC-K, MDP, MRT, AM and PF contributed to the design, theoretical analysis and implementation of the analysis pipeline. SRS, MDP, JC-K, NSV, LT, CH, TMJ, CJM, PJS, NB, EP and ESAS were responsible for data acquisition. SRS, MDP, JC-K, NSV, JK and PJS were responsible for analysis of data. JK, PJS, MDP and JC-K were responsible for interpretation of data. PJS, JK and MDP supervised the project. The Project MinE ALS Sequencing Consortium (online supplementary table 4) was involved in data acquisition and analysis. All authors meet the four ICMJE authorship criteria, and were responsible for revising the manuscript, approving the final version for publication, and for accuracy and integrity of the work.

Funding PJS is supported as an NIHR Senior Investigator (NF-SI-0617–10077). PJS, SRS, NV, JK, PF, MRT and AM are supported by the Motor Neurone Disease Association (AMBRoSIA 972-797 and NECTAR 974-797 awards). LT is supported by the NIHR Yorkshire and Humber Clinical Research Network. This research was supported by the NIHR Sheffield Biomedical Research Centre (BRC) (IS-BRC-1215-20017) and the NIHR Sheffield Clinical Research Facility (CRF).

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. All data relevant to the study are included in the article or uploaded as online supplemental information. Data are stored in the AMBRoSIA consortium repository. Anonymised raw data can be made available on reasonable request.

ORCID iDs
Stephanie R Shepheard http://orcid.org/0000-0002-2358-7348
Johnathan Cooper-Knock http://orcid.org/0000-0002-0873-8689
Martin R Turner http://orcid.org/0000-0003-0267-3180
Thomas M Jenkins http://orcid.org/0000-0003-0342-7184
Christopher J McDermott http://orcid.org/0000-0002-1269-9053
Pamela J Shaw http://orcid.org/0000-0002-8925-2567

REFERENCES

1 Strong MJ, Grace GM, Freedman M, *et al*. Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* 2009;10:131–46.
2 Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 2014;17:17–23.
3 Cooper-Knock J, Hewitt C, Highley JR, *et al*. Clinico-pathological features in amyotrophic lateral sclerosis with expansions in C9ORF72. *Brain* 2012;135:751–64.
4 Renton AE, Majounie E, Waite A, *et al*. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011;72:257–68.
5 Müller K, Brenner D, Weydt P, *et al*. Comprehensive analysis of the mutation spectrum in 301 German ALS families. *J Neurol Neurosurg Psychiatry* 2018;89:817–27.
6 Rosen DR, Siddique T, Patterson D, *et al*. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 1993;362:59–62.
7 Kabashi E, Valdmanis PN, Dion P, *et al*. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet* 2008;40:572–4.
8 Sreedharan J, Blair IP, Tripathi VB, *et al*. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* 2008;319:1668–72.
9 Vance C, Rogelj B, Hortobágyi T, *et al*. Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* 2009;323:1208–11.
10 Cady J, Allred P, Bali T, *et al*. Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Ann Neurol* 2015;77:100–13.
11 McCann EP, Henden L, Fifita JA, *et al*. Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis. *J Med Genet* 2020:jmedgenet-2020-106866.
12 Al-Chalabi A, Calvo A, Chio A, *et al*. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. *Lancet Neurol* 2014;13:1108–13.
13 Chio A, Logroscino G, Hardiman O, *et al*. Prognostic factors in ALS: a critical review. *Amyotrophic Lateral Sclerosis* 2008:1–14.
14 Turner MR, Al-Chalabi A, Chio A, *et al*. Genetic screening in sporadic ALS and FTD. *J Neurol Neurosurg Psychiatry* 2017;88:1042–4.
15 Miller RG, Mitchell JD, Moore DH. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane Database Syst Rev* 2012:CD001447.
16 Yoshino H, Kimura A. Investigation of the therapeutic effects of edaravone, a free radical scavenger, on amyotrophic lateral sclerosis (phase II study). *Amyotrophic Lateral Sclerosis* 2006;7:247–51.
17 Malone ER, Oliva M, Sabatini PJB, *et al*. Molecular profiling for precision cancer therapies. *Genome Med* 2020;12:8.
18 Sproviero W, Shatunov A, Stahl D, *et al*. ATXN2 trinucleotide repeat length correlates with risk of ALS. *Neurobiol Aging* 2017;51:178.e1–9.
19 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
20 McKenna A, Hanna M, Banks E, *et al*. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.
21 Tan A, Abecasis GR, Kang HM. Unified representation of genetic variants. *Bioinformatics* 2015;31:2202–4.
22 Karczewski KJ, Francioli LC, Tiao G, *et al*. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *BioRxiv* 2019.
23 Schwarz JM, Rödelsperger C, Schuelke M, *et al*. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* 2010;7:575–6.
24 Adzhubei IA, Schmidt S, Peshkin L, *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
25 Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res* 2003;31:3812–4.
26 Cooper GM, Stone EA, Asimenos G, *et al*. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* 2005;15:901–13.
27 Rentzsch P, Witten D, Cooper GM, *et al*. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res* 2019;47:D886–94.
28 Reese MG, Eeckman FH, Kulp D, *et al*. Improved splice site detection in genie. *J Comput Biol* 1997;4:311–23.

29 Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001;29:1185–90.
30 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* 2004;11:377–94.
31 Landrum MJ, Lee JM, Riley GR, *et al*. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014;42:D980–5.
32 Ellard S, Lindsay H, Camm N, *et al*. Practice guidelines for targeted next generation sequencing analysis and interpretation. *ACGS Best Practice Guidelines* 2014.
33 Osmanovic A, Rangnau I, Kosfeld A, *et al*. FIG4 variants in central European patients with amyotrophic lateral sclerosis: a whole-exome and targeted sequencing study. *Eur J Hum Genet* 2017;25:324–31.
34 Yang Y, Hentati A, Deng HX, *et al*. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nat Genet* 2001;29:160–5.
35 Project MinE ALS Sequencing Consortium. Project mine: study design and pilot analyses of a large-scale whole-genome sequencing study in amyotrophic lateral sclerosis. *Eur J Hum Genet* 2018;26:1537–46.
36 Freischmidt A, Wieland T, Richter B, *et al*. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat Neurosci* 2015;18:631–6.
37 Cooper-Knock J, Robins H, Niedermoser I, *et al*. Targeted genetic screen in amyotrophic lateral sclerosis reveals novel genetic variants with synergistic effect on clinical phenotype. *Front Mol Neurosci* 2017;10:370.
38 Pesiridis GS, Lee VM-Y, Trojanowski JQ. Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Hum Mol Genet* 2009;18:R156–62.
39 Kenna KP, van Doormaal PTC, Dekker AM, *et al*. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. *Nat Genet* 2016;48:1037–42.
40 van Blitterswijk M, van Es MA, Hennekam EAM, *et al*. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet* 2012;21:3776–84.