Original research

Evidence for polygenic and oligogenic basis of Australian sporadic amyotrophic lateral sclerosis

Emily P McCann [iD],[1] Lyndal Henden [iD],[1] Jennifer A Fifita [iD],[1] Katharine Y Zhang,[1] Natalie Grima [iD],[1] Denis C Bauer [iD],[2] Sandrine Chan Moi Fat [iD],[1] Natalie A Twine [iD],[1,2] Roger Pamphlett [iD],[3,4,5] Matthew C Kiernan,[4,6] Dominic B Rowe [iD],[1,7] Kelly L Williams [iD],[1] Ian P Blair [iD][1]

ABSTRACT

Background Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease with phenotypic and genetic heterogeneity. Approximately 10% of cases are familial, while remaining cases are classified as sporadic. To date, >30 genes and several hundred genetic variants have been implicated in ALS.

Methods Seven hundred and fifty-seven sporadic ALS cases were recruited from Australian neurology clinics. Detailed clinical data and whole genome sequencing (WGS) data were available from 567 and 616 cases, respectively, of which 426 cases had both datasets available. As part of a comprehensive genetic analysis, 853 genetic variants previously reported as ALS-linked mutations or disease-associated alleles were interrogated in sporadic ALS WGS data. Statistical analyses were performed to identify correlation between clinical variables, and between phenotype and the number of ALS-implicated variants carried by an individual. Relatedness between individuals carrying identical variants was assessed using identity-by-descent analysis.

Results Forty-three ALS-implicated variants from 18 genes, including C9orf72, ATXN2, TARDBP, SOD1, SQSTM1 and SETX, were identified in Australian sporadic ALS cases. One-third of cases carried at least one variant and 6.82% carried two or more variants, implicating a potential oligogenic or polygenic basis of ALS. Relatedness was detected between two sporadic ALS cases carrying a SOD1 p.I114T mutation, and among three cases carrying a SQSTM1 p.K238E mutation. Oligogenic/polygenic sporadic ALS cases showed earlier age of onset than those with no reported variant.

Conclusion We confirm phenotypic associations among ALS cases, and highlight the contribution of genetic variation to all forms of ALS.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a late-onset fatal neurodegenerative disease characterised by the degeneration of both upper and lower motor neurons.[1] Patients develop muscle weakness, wasting, spasticity and eventual paralysis. ALS displays significant phenotypic heterogeneity, even among cases with identical causal gene mutations. Disease onset can occur anywhere from the second to ninth decade of life but most frequently between 50 and 60 years of age.[2] Around 70% of cases present with limb onset and 25% with bulbar onset, while rare cases show truncal onset.[3] The median disease course is 3 years,[2] however survival can be as short as 12 months and may exceed 20 years.[2 4] Around 10%–15% of ALS cases are diagnosed with comorbid frontotemporal dementia (FTD), with up to 50% developing some cognitive impairment.[5] Approximately 10% of ALS cases have a family history of disease (familial ALS (FALS)) and two-thirds of these cases carry a reported ALS gene mutation.[6 7] The remaining 90% of cases have no apparent family history and are classified as sporadic ALS (SALS).

Extensive genetic heterogeneity is apparent among ALS cases. To date, at least 31 genes have been linked or associated with ALS. Heritability studies suggest that 40%–60% of SALS risk may be explained by genetic factors.[8–10] A polygenic basis to SALS has been implicated by the co-occurrence of two or more ALS gene variants in a single individual.[11–14] These variants likely interact with environmental factors to trigger the development of ALS.[15] Alternatively, this may indicate an oligogenic disease model,[11] where an ALS mutation of large effect, such as SOD1 or C9orf72, is inherited with another ALS gene variant which may also contribute to phenotype. Furthermore, a multistep hypothesis has recently been postulated to explain the late and apparently sporadic onset of ALS.[16] This hypothesis postulates that six 'steps' are required to trigger ALS onset, where such steps may include genetic predisposition, environmental exposures or other unknown molecular alterations.[16] Within this hypothesis, known genetic mutations may account for multiple steps, to a degree reflective of their penetrance.[17]

Previously, our laboratory described the genetic and phenotypic heterogeneity of Australian familial ALS.[7] Here, we report the extent of phenotypic and genetic variation among Australian SALS cases. To interrogate the genetic architecture of Australian SALS, we first compiled a list of 31 genes, and >850 genetic variants previously reported as: 1) ALS-linked mutations (typically identified from family based studies); 2) functional risk alleles (variants whose functions are thought to increase disease susceptibility) or 3) otherwise associated variants (including variants that may be in linkage disequilibrium with unknown functional risk alleles).

Among a large Australian SALS cohort, 35.39% of cases carried a variant previously reported in ALS, implicating 18 genes as contributing to disease in this population. Polygenic inheritance was implicated in 6.82% of SALS cases who harboured more than one ALS-implicated variant. We also showed that the clinical heterogeneity of Australian SALS is similar to that observed in cohorts from other populations.

METHODS
Subjects
A total of 757 SALS cases were recruited from the Macquarie University Neurodegenerative Disease Biobank, Australian MND DNA Bank (Royal Prince Alfred Hospital) and Brain and Mind Centre (University of Sydney). All participants provided informed written consent as approved by the human research ethics committees of Macquarie University, Sydney South West Area Health District or The University of Sydney. All participants were of Caucasian descent (as established according to online supplementary file 1), each was clinically diagnosed with definite or probable ALS according to *El Escorial* criteria,[18] and had no known relatives affected by ALS and/or FTD. Of these 757 cases, detailed clinical data were available for 567 individuals, and 616 cases underwent whole genome sequencing (WGS), with a total of 426 having both detailed clinical and WGS data available. Genomic DNA was extracted from peripheral blood using standard protocols. Fifteen of these SALS cases were previously reported to carry an expansion in *C9orf72*.[7 14]

Statistical analysis of clinical variables
Clinical records were examined for four phenotypic features: sex, age at disease onset, site of onset (bulbar or spinal) and duration of disease from onset (until death or last known date of survival). All statistical analyses were performed in R (V.3.5.1). Statistical analyses were performed in a pairwise fashion between all four clinical variables to identify significant associations. A χ^2 analysis was performed between sex and site of onset, while Welch's t-tests were performed between age of onset and both sex and site of onset. Kaplan-Meier survival analyses were performed between disease duration and both sex and site of onset. Additionally, a linear regression model was fitted between age of onset and duration (for deceased cases only). Multiple testing was accounted for using a Bonferroni corrected significance threshold of $p < 0.008$, with $\alpha = 0.05$ and 6 comparisons.

We also assessed whether the number of ALS-implicated variants carried by an individual influenced their clinical presentation. Multinomial logistic regression analysis was performed for each clinical variable to compare cases carrying two or more ALS-implicated variants with cases carrying only one ALS-implicated variant or no ALS-implicated variants.

Generation of WGS data
DNA samples underwent library preparation using the TruSeq PCR-free library preparation kit (Illumina, V.2.5). Prepared libraries then underwent multiplex 150 bp paired-end sequencing using an Illumina HiSeq X Ten instrument (Kinghorn Centre for Clinical Genomics, Sydney, Australia). Raw sequencing reads were processed using the genome analysis toolkit and the associated best practices.[19–21] Detailed methodology can be found in online supplementary file 1.

Survey of reported genetic variants implicated in ALS
A list of 31 ALS genes was established (online supplementary table 1) based on the 28 ALS genes reported by Chia *et al*,[22]

with the addition of the recently reported ALS genes *TIA1*[23] and *KIF5A*,[24] as well as the *GPX3-TNIP1* locus.[25] A comprehensive literature search was conducted in order to compile an exhaustive reference list of 853 genetic variants previously implicated in ALS (online supplementary file 2), including ALS-linked mutations, functional risk alleles and disease-associated variants. This involved performing a PubMed (https://www.ncbi.nlm.nih.gov/pubmed/) search for each ALS gene name and 'amyotrophic lateral sclerosis', with subsequent manual evaluation of all resulting publications. The ALSOD database (http://alsod.iop.kcl.ac.uk/home.aspx; last updated 8 September 2015)[26] was also consulted.

Genetic variants were included in this reference list (online supplementary file 2) if they were predicted to alter the amino acid sequence of a protein (ie, non-synonymous/missense, frameshift and non-frameshift insertions/deletions, splicing variants) and were named in the main text of the publication (ie, variants included in an aggregate of variants used for burden testing were not added to the list). Details of genomic location (hg19), the transcript accession number and cDNA and protein changes were recorded for each variant. For variants where these details were not reported in the original publication(s), they were determined using the University of California Santa Cruz Variant Annotation Integrator (http://genome.ucsc.edu/cgi-bin/hgVai) and the associated Human Genome Variation Society variant nomenclature track. Ancillary details pertaining to the ancestry and ALS inheritance mode for cases carrying each variant were also recorded, as was the presence of each variant in unaffected and unrelated control individuals, where available in the original publication(s).

Identifying ALS-implicated variants in SALS cases
WGS data from 616 SALS cases were parsed for 31 ALS genes (online supplementary table 1) using custom UNIX scripts. Filtering using R was subsequently used to identify the previously reported ALS-implicated variants (online supplementary file 2) present within this dataset. R scripts were also used to identify cases that carried each variant and, conversely, which variants were carried by each case. Unless otherwise specified, all minor allele frequency (MAF) values used for comparisons were from the ethnically matched non-neuro (ie, excluding individuals with neurological disorders) non-Finnish European (NFE) subset of the Genome Aggregation Database (gnomAD; n=51 592).[27]

C9orf72 GGGGCC and *ATXN2* CAG repeat genotyping was performed on WGS BAM files using ExpansionHunter[28 29] with default settings. These genotypes were interrogated and combined with the VCF data obtained above using R. Validation was performed for cases with expanded (>30 repeats) or intermediate (23–30 repeats) *C9orf72* repeat lengths using repeat primed PCR,[30] and those with intermediate *ATXN2* repeat lengths (29–39 repeats) using conventional PCR.[31] PCR products were analysed by fragment length analysis using an ABI 3730XL DNA Analyser (Applied Biosystems), and size analysis was performed using GeneMarker (V.3.0.1) software (SoftGenetics, Pennsylvania, USA).

Association analysis
To determine whether any of the reported ALS-implicated variants were associated with SALS in Australia, allele counts were compared between SALS cases (n=616) and control individuals from the non-neuro NFE subset of gnomAD (n=51 592)[27] using Fisher's exact test in R. A Bonferroni corrected significance threshold of $p < 5.875 \times 10^{-5}$ was applied to account for the

851 nucleotide level variants tested for association with disease. Control genotyping data for repeat expansions in *C9orf72* and *ATXN2* were not available, therefore association testing was not completed for these variants. In cases where an ALS-implicated variant was not reported in gnomAD, allele counts were determined by inferring the number of gnomAD controls covered at the site of interest, based on the number of individuals with genotypes for variants located within 40 bp. Data from SALS cases (n=4366) and control individuals (n=1832) from Project MiNE[32] were used for validation purposes.

Identity-by-descent analysis in cases carrying identical ALS-implicated variants

Identity-by-descent (IBD) analysis was performed to investigate whether SALS cases that carried an identical ALS-implicated variant had inherited the variant from a recent common ancestor. Relatedness analysis was performed on WGS data using TRIBES with default parameter settings.[33] Briefly, TRIBES filtered the WGS data according to quality control metrics, phased the resulting genotype calls and performed pairwise IBD inference. IBD segments were then examined to determine if SALS cases that carried an identical ALS-implicated variant were inferred IBD over that locus. Graphical networks were then produced using isoRelate[34] to show putatively related cases that carried identical variants inherited from a common ancestor.

RESULTS
Phenotypic variability of Australian SALS was consistent with that of European cohorts

The clinical dataset of 567 SALS cases comprised 61% males and 71% spinal onset cases. The average age of onset was 60 years (SD=12 years, Q1=54 years, median=62 years, Q3=69 years), and the average disease duration was 38 months (SD=19 months, Q1=26 months, median=35 months, Q3=47 months), at which time 62% of cases were deceased.

Statistical analyses were performed to identify any association between clinical variables including sex, age at disease onset, site of disease onset and disease duration (figure 1). A significant association was observed between sex and site of onset (p=0.00015, figure 1A), where females were more likely to develop bulbar onset while more males presented with spinal onset. Females also had a significantly shorter life expectancy than males (p=0.0079, figure 1C). While females tended to have a later age of onset (figure 1B) this result was not statistically significant. There was also evidence of a relationship between age of onset and disease duration (p=0.00092, figure 1D), with shorter disease duration in cases with a later onset. Cases with bulbar onset were more likely to have a later age of onset (p=0.00031, figure 1E) and reduced life expectancy (p<0.0001, figure 1F).
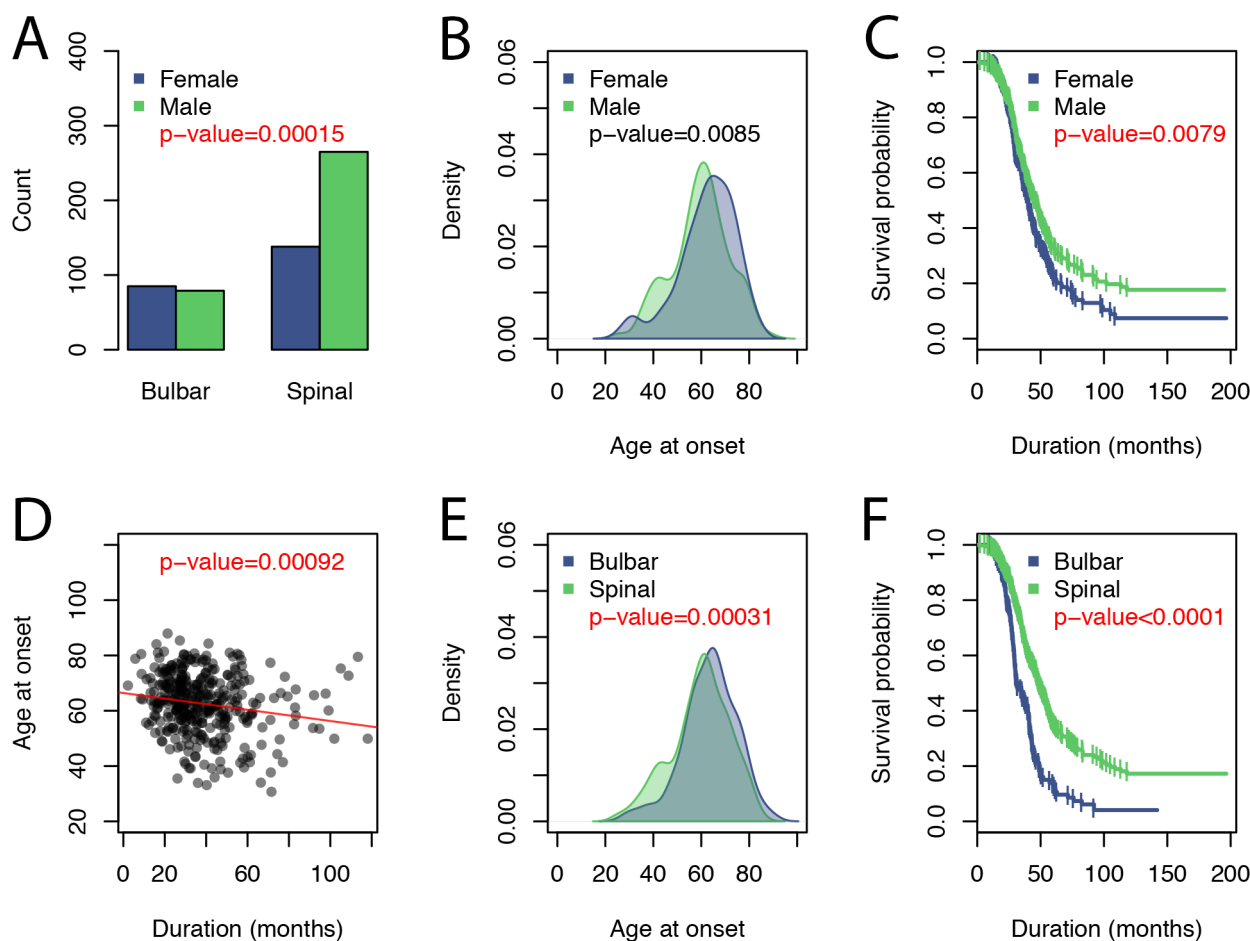
Figure 1 Statistical analyses of clinical variables for 573 sporadic amyotrophic lateral sclerosis (SALS) cases. P values highlighted in red are significant at the Bonferroni-adjusted significance level p=0.008. There was a significant difference in site of onset between males and females (A), where females presented with bulbar onset more often. Females had a later age of onset (B) and a significantly shorter life expectancy (C). There was a significant association between age of onset and disease duration (D) and cases with bulbar onset had a later age of onset (E) as well as a shorter life expectancy (F).

More than 30% of Australian SALS cases carried a genetic variant previously implicated in ALS

We generated a comprehensive list of 853 genetic variants from 31 genes that were previously implicated in ALS by peer-reviewed publications between 1993 and February 2019 (online

supplementary file 2). This included repeat expansions in *C9orf72* and *ATXN2*, as well as 851 single nucleotide and insertion/deletion variants in 29 additional genes. Of the 853 variants implicated in ALS, 46 variants (from 20 genes) were identified across the cohort of 616 Australian SALS cases (table 1). This

Table 1	ALS-implicated variants identified among 616 Australian SALS cases					
			Frequency in SALS (n=616)		Association analysis (Fisher's exact test p value)	
Gene name	Nucleotide alteration	Consequence	No. heterozygous cases	No. homozygous cases	SALS vs gnomAD non-neuro NFE	Project MiNE cases vs controls
ANG	c.A122T	p.K41I	2	0	1	.
ANG	c.A208G	p.I70V	1	0	1	.
ATXN2	CAG expansion (29–39 repeats)		10	0	N/A	N/A
C21orf2	c.G172T	p.V58L	22	1	0.0140124	.
C9orf72	GGGGCC repeat expansion (>30 repeats)		41	0	N/A	N/A
CCNF	c.T1810A	p.F604I	1	0	1	1
CCNF	c.G2140A	p.V714M	17	1	0.8111191	.
CHCHD10	c.C100T	p.P34S	6	0	0.6427256	.
CHCHD10	c.C239T	p.P80L	2	0	0.0377862	.
CHCHD10	c.T403C	p.Y135H	1	0	0.3094276	.
DCTN1	c.C3746T	p.T1249I	6	0	1	.
*ELP3**	g.28086088G>A	Unknown	227	345	0.0433814	.
*ELP3**	g.28136109T>C	Unknown	104	502	0.5108734	.
FUS	c.*41G>A	Unknown	12	0	0.4211107	.
FUS	c.833–29C>T	Unknown	15	1	0.9006159	.
*GPX3-TNIP1**	c.*3144C>T	Unknown	248	315	0.0015471	.
NEFH	c.2230_2247del	Unknown	1	0	0.3617984	.
NEK1	c.G782A	p.R261H	3	0	0.6387112	.
NEK1	**c.C3107G**	**p.S1036X**	**5**	**0**	**0.0000102**	**0.000647562**
OPTN	c.T293A	p.M98K	26	2	0.4873868	.
OPTN	c.G403T	p.E135X	2	0	0.0188011	.
OPTN	c.G476T	p.G159V	1	0	0.0401598	.
SETX	c.A431G	p.N144S	1	0	0.0234651	.
SETX	c.G2755C	p.V919L	1	0	0.1432519	.
SETX	c.A2975G	p.K992R	10	0	0.0065459	.
SETX	c.C4433A	p.A1478E	1	0	0.4917794	.
SETX	c.T4660G	p.C1554G	1	0	0.1388244	.
SETX	c.T7640C	p.I2547T	8	0	0.7054959	.
SOD1	c.A272C	p.D91A	1	0	0.5323133	.
SOD1	**c.T341C**	**p.I114T**	**3**	**0**	**0.0000099**	**0.189084762**
SPG11	c.C491T	p.S164L	0	1	0.0071392	.
SPG11	c.A6224G	p.N2075S	5	0	1	.
SQSTM1	c.C98T	p.A33V	1	0	1	.
SQSTM1	c.A712G	p.K238E	7	0	0.2207281	.
SQSTM1	c.G822C	p.E274D	30	0	0.699596	.
SQSTM1	c.C1175T	p.P392L	3	0	0.2471604	.
SQSTM1	g.3'+7G>C	unknown	1	0	0.1246189	.
SQSTM1	g.5'–37C>T	unknown	1	0	0.510332	.
TARDBP	c.543+112C>A	unknown	2	0	1	.
TARDBP	c.G859A	p.G287S	1	0	0.0401616	.
TARDBP	c.G883A	p.G295S	1	0	0.0135711	.
TARDBP	c.G1144A	p.A382T	1	0	0.0292553	.
TARDBP	c.A1147G	p.I383V	2	0	0.000216	.
TBK1	c.A871G	p.K291E	1	0	0.2670662	.
TBK1	c.G1073A	p.R358H	1	0	0.0350117	.
UBQLN2	c.G1019T	p.S340I	1	0	0.0557581	.

Variants that were significantly associated with Australian SALS are indicated in bold.

*These variants had MAF >0.2 among gnomAD non-neuro NFE controls and were removed from further analysis as common variants.

gnomAD, Genome Aggregation Database; MAF, minor allele frequency; N/A, not available; NFE, non-Finnish European; SALS, sporadic amyotrophic lateral sclerosis.

included three common population-based (MAF >0.2) intronic SNPs, *ELP3* rs2614046 and rs6985069, and *GPX3-TNIP1* rs10463311. Given their high frequency in both cases and controls, independent of their potential contribution to disease, they were not considered in any further analyses. After removal of these common SNPs, 43 ALS-implicated variants from 18 different genes were present among 218/616 (35.39%) SALS cases, the majority of which were heterozygous. This included 41 cases that carried a *C9orf72* hexanucleotide repeat expansion (15 of which had been previously identified using repeat-primed PCR[7 14]) and 10 cases that harboured intermediate-sized repeat expansions in *ATXN2*. Three additional cases carried intermediate-sized hexanucleotide repeat expansions in *C9orf72*, although were not considered *C9orf72* expansion positive. The FALS-linked *SOD1* p.I114T and the *SOD1* p.D91A mutations were identified in three and one Australian SALS case(s), respectively. Missense *TARDBP* mutations p.G287S, p.G295S and p.A382T were each identified in a single SALS case and *TARDBP* p.I383V was present in two SALS cases. The remaining ALS-implicated variants identified in Australian SALS cases had more limited evidence to support their role in ALS and were observed in gnomAD non-neuro NFE controls. The relatively common variants (MAF >0.01), *CCNF* p.V714M, *C21orf2* p.V58L, *OPTN* p.M98K and *FUS* c.833–29C>T, were each observed as both heterozygous and homozygous variants, and *SPG11* p.S164L was exclusively observed in a homozygous state.

NEK1 and *SOD1* mutations showed significant association with Australian SALS

Fisher's exact test determined that two variants, *NEK1* p.S1036X ($p=1.020\times10^{-5}$) and *SOD1* p.I114T ($p=9.871\times10^{-6}$) were significantly over-represented in Australian SALS cases compared with gnomAD non-neuro NFE controls. A further 14 variants were over-represented among cases with nominally significant p values ($5.875\times10^{-5}<p<0.05$) (table 1, online supplementary table 2). *NEK1* p.S1036X showed nominal association with disease ($p=6.47\times10^{-4}$) in the Project MiNE case-control cohort, although did not reach corrected significance ($p<5.875\times10^{-5}$, table 1, online supplementary table 2).

Some Australian SALS cases who carried identical ALS-implicated variants may be distantly related

Of the 23 ALS-implicated variants that were identified in multiple SALS cases (table 1), IBD segments and an estimated degree of relatedness were determined between 18 pairs of cases over 8 different ALS-implicated variants (online supplementary table 3, online supplementary figure 1). Other than a pair of *SOD1* p.I114T positive cases, these IBD segments ranged between 3 and 12 cM, corresponding to degree of relatedness between 8th and 11th degree. Notably, three trios shared IBD segments, one with each of *SQSTM1* p.K238E (population MAF=0.003558, average IBD segment 9.08 cM), *CCNF* p.V714M (population MAF=0.01469, average IBD segment 4.57 cM) and *DCTN1* p.T1249I (population MAF=0.004954, average IBD segment 3.49 cM). No IBD segments >3 cM were detected between the 41 SALS cases with a *C9orf72* hexanucleotide repeat expansion, nor 10 cases with intermediate-sized *ATXN2* expansions, or the two SALS cases who carried *TARDBP* p.I383V. As part of an extended IBD analysis, the three cases who carried *SOD1* p.I114T were linked as sixth degree relatives to existing Australian FALS families, therefore re-classifying these apparently sporadic cases as misclassified FALS.[35]

Potential oligogenic/polygenic basis of ALS in Australian SALS

A total of 42/616 (6.82%) Australian SALS cases were found to harbour more than one reported ALS-implicated variant (table 2) (after excluding the three common variants found in this cohort). Of these 42 individuals, 38 carried two ALS-implicated variants and 4 carried three variants. The cases who carried additional ALS-implicated variants together with *C9orf72* (n=17) or *SOD1* (n=3) mutations may represent oligogenic inheritance. Notably, all three individuals who carried the known FALS-linked *SOD1* p.I114T mutation carried at least one other ALS-implicated variant. Potential polygenic inheritance is implicated for all other cases who carried multiple disease-associated variants (with population MAF values between 0.000022 and 0.02815). *SQSTM1* p.E274D and *OPTN* p.M98K were identified in 11 and 10 cases, respectively, including in an oligogenic state in 5 cases carrying a *C9orf72* expansion.

Significant association between putative polygenic/oligogenic SALS cases and the age at disease onset

Multinomial logistic regression analysis showed that cases with a younger age of onset were more likely to carry multiple ALS-implicated variants (OR 0.964, p=0.025; online supplementary table 4). No other clinical variables were significantly associated with polygenic/oligogenic variation.

DISCUSSION

We sought to characterise the phenotypic and genetic heterogeneity of sporadic ALS by surveying clinical characteristics and the presence of sequence variants previously reported to be pathogenic, or associated with ALS, among a large cohort of Australian SALS cases. We also sought to determine whether cases harbouring multiple ALS-implicated variants exhibited more severe clinical characteristics. We demonstrated a high degree of genetic heterogeneity among Australian SALS, with 43 different variants from 18 ALS genes identified among 35.39% of cases. We showed that 6.82% of Australian SALS potentially have a polygenic or oligogenic underpinning to disease, and that patient clinical phenotype was likely influenced by the presence of more than one ALS-implicated variant. Other significant associations that were found between clinical variables in our Australian cohort were consistent with reported findings and confirm the phenotypic heterogeneity of SALS.[36 37] This work highlights the genetic contribution and heterogeneity of SALS and suggests that polygenic and/or oligogenic mechanisms may be at play.

Our literature search revealed that 853 genetic variants have been implicated in ALS over the past 26 years. Of these, 43 variants from 18 genes were identified among 35.39% of Australian SALS. As these sporadic cases have no known family history of disease, this high prevalence of reported ALS-implicated variants may reflect unrecognised false positive variants in the literature. Indeed, the evidence supporting the pathogenicity of each ALS-implicated variant varies significantly. While many have strong support as pathogenic ALS mutations through genetic linkage studies and/or extensive segregation within families, others have less compelling evidence to support their role in ALS pathogenesis. For example, variants identified in single FALS or SALS cases, or very small ALS families, have been reported as pathogenic on the basis that they fall within an established ALS gene despite a lack of supporting segregation data. In addition, candidate variants identified prior to the widespread adoption of high-throughput sequencing technologies were typically screened through <100 control individuals. Now that large-scale control databases are available with next-generation sequencing data,

Table 2 Summary of SALS cases who carried multiple ALS-implicated variants

Sample name	Variant 1	Variant 2	Variant 3	Polygenic or oligogenic inheritance?
MQ140094	*ATXN2* CAG repeat expansion	*SETX* c.T7640C, p.I2547T	.	Polygenic
SALS0890	*ATXN2* CAG repeat expansion	*SQSTM1* c.G822C, p.E274D	.	Polygenic
SALS1278	*ATXN2* CAG repeat expansion	*SQSTM1* c.G822C, p.E274D	.	Polygenic
SALS0396	*C21orf2* c.G172T, p.V58L	*SPG11* c.A6224G, p.N2075S	.	Polygenic
SALS1724	*C9orf72* GGGGCC expansion	*ATXN2* CAG repeat expansion	*NEK1* c.G782A, p.R261H	Oligogenic
SALS0782	*C9orf72* GGGGCC expansion	*C21orf2* c.G172T, p.V58L	.	Oligogenic
SALS1806	*C9orf72* GGGGCC expansion	*C21orf2* c.G172T, p.V58L	.	Oligogenic
MQ140100	*C9orf72* GGGGCC expansion	*CHCHD10* c.C239T, p.P80L	.	Oligogenic
SALS0846	*C9orf72* GGGGCC expansion	*FUS* c.*41G>A, intronic	.	Oligogenic
SALS1354	*C9orf72* GGGGCC expansion	*FUS* c.*41G>A, intronic	.	Oligogenic
SALS1574	*C9orf72* GGGGCC expansion	*FUS* c.833–29C>T, intronic	.	Oligogenic
SALS1488	*C9orf72* GGGGCC expansion	*FUS* c.833–29C>T, intronic	*OPTN* c.T293A, p.M98K	Oligogenic
SALS2337	*C9orf72* GGGGCC expansion	*OPTN* c.G476T, p.G159V	.	Oligogenic
SALS0859	*C9orf72* GGGGCC expansion	*OPTN* c.T293A, p.M98K	.	Oligogenic
SALS1387	*C9orf72* GGGGCC expansion	*SETX* c.T7640C, p.I2547T	.	Oligogenic
SALS0326	*C9orf72* GGGGCC expansion	*SPG11* c.A6224G, p.N2075S	.	Oligogenic
SALS1090	*C9orf72* GGGGCC expansion	*SPG11* c.A6224G, p.N2075S	.	Oligogenic
SALS0934	*C9orf72* GGGGCC expansion	*SQSTM1* c.G822C, p.E274D	.	Oligogenic
SALS1522	*C9orf72* GGGGCC expansion	*SQSTM1* c.G822C, p.E274D	.	Oligogenic
SALS1960	*C9orf72* GGGGCC expansion	*SQSTM1* c.G822C, p.E274D	.	Oligogenic
SALS2359	*C9orf72* GGGGCC expansion	*SQSTM1* c.G822C, p.E274D	.	Oligogenic
SALS1910	*CHCHD10* c.C100T, p.P34S	*SETX* c.T7640C, p.I2547T	.	Polygenic
SALS1980	*CHCHD10* c.C100T, p.P34S	*SQSTM1* c.G822C, p.E274D	.	Polygenic
MQ130086	*FUS* c.*41G>A, intronic	*CCNF* c.G2140A, p.V714M	.	Polygenic
SALS1809	*FUS* c.*41G>A, intronic	*OPTN* c.G403T, p.E135X	.	Polygenic
MQ150164	*FUS* c.*41G>A, intronic	*OPTN* c.T293A, p.M98K	.	Polygenic
SALS0226	*FUS* c.*41G>A, intronic	*OPTN* c.T293A, p.M98K	*CHCHD10* c.C239T, p.P80L	Polygenic
MQ140255	*FUS* c.833–29C>T, intronic	*NEK1* c.C3107G, p.S1036X	.	Polygenic
MN201410	*FUS* c.833–29C>T, intronic	*OPTN* c.T293A, p.M98K	.	Polygenic
SALS0312	*FUS* c.833–29C>T, intronic	*SETX* c.A2975G, p.K992R	.	Polygenic
SALS2282	*NEK1* c.C3107G, p.S1036X	*ANG* c.A122T, p.K41I	.	Polygenic
MQ160055	*NEK1* c.G782A, p.R261H	*OPTN* c.T293A, p.M98K	.	Polygenic
SALS1700	*OPTN* c.T293A, p.M98K	*NEFH* c.2230_2247del, intronic	.	Polygenic
SALS2258*	*SOD1* c.T341C, p.I114T	*CCNF* c.G2140A, p.V714M	*C21orf2* c.G172T, p.V58L	Oligogenic
MN201517*	*SOD1* c.T341C, p.I114T	*SETX* c.A2975G, p.K992R	.	Oligogenic
SALS1259*	*SOD1* c.T341C, p.I114T	*SQSTM1* c.G822C, p.E274D	.	Oligogenic
MQ140090	*SQSTM1* c.G822C, p.E274D	*ANG* c.A208G, p.I70V	.	Polygenic
SALS1380	*SQSTM1* c.G822C, p.E274D	*DCTN1* c.C3746T, p.T1249I	.	Polygenic
SALS0189	*SQSTM1* c.G822C, p.E274D	*OPTN* c.T293A, p.M98K	.	Polygenic
SALS2285	*SQSTM1* g.3'+7G>C, intronic	*OPTN* c.T293A, p.M98K	.	Polygenic
SALS1130	*TARDBP* c.543+112C>A, intronic	*C21orf2* c.G172T, p.V58L	.	Polygenic
MQ140199	*TBK1* c.A871G, p.K291E	*OPTN* c.T293A, p.M98K	.	Polygenic

*Reclassified as FALS in the study by Henden *et al*.[35]
FALS, familial amyotrophic lateral sclerosis; SALS, sporadic amyotrophic lateral sclerosis.

it has become apparent that some reported putative ALS mutations are instead rare population-based variants, such as *SOD1* p.N20S (online supplementary file 2). Beyond the variants analysed in the present study, it is important to note that unreported rare and novel genetic variants in the known ALS genes may also contribute to the development and phenotypic presentation of SALS.

Twelve ALS genes were found to harbour more than one unique variant in our cohort. Most notably, this included five variants in the known ALS gene, *TARDBP*. Four of these *TARDBP* missense mutations (p.G287S, p.G295S, p.A382T and p.I383V) have also been reported in additional SALS cohorts of various origins (online supplementary file 2) and are extremely rare in the healthy population (MAF <0.00002233), observations that support the pathogenic role of these variants. It is likely that these variants are low penetrance mutations, given their absence from FALS cohorts as established by the literature search conducted here. Alternatively, these may be de novo variants within mutation hot spots. Another possibility is they may be risk alleles, which is supported by our association analyses where all four *TARDBP* missense mutations showed nominal association with SALS (table 1). IBD analysis failed to detect a relationship between two Australian SALS cases who carried an identical *TARDBP* mutation (p.I383V) suggesting that it is either an old mutation or arose independently more recently.

An interesting identified variant was the homozygous *SPG11* p.S164L. The original report of this rare variant also identified a homozygous SALS case,[38] although it was suggested this variant may be a benign polymorphism.[38] It is interesting to note that no individuals in the gnomAD non-neuro control cohort (of any ancestry) are homozygous carriers of this variant. It is possible that homozygous *SPG11* p.S164L confers increased susceptibility to ALS through loss of-function.

Association testing identified just two variants that were significantly over-represented among Australian SALS cases when compared with controls. One of these was the FALS causal *SOD1* p.I114T variant, although the putative SALS cases that carried this variant were subsequently found to be unrecognised FALS cases.[35] The other over-represented variant, *NEK1* p.S1036X, was previously reported in a single FALS case, although segregation was not demonstrated in the family.[39] Together with our data, this suggests that *NEK1* p.S1036X is an ALS susceptibility allele of moderate-to-high effect, with implications for other *NEK1* loss-of-function variants.

We sought to determine whether SALS cases that shared the same rare ALS-implicated variants (MAF <0.03 in the general population) were distantly related and potentially represented unrecognised FALS cases. IBD analysis demonstrated that the trio of SALS cases that carried *SQSTM1* p.K238E were likely to be distantly related, with shared IBD segments of at least 7.892 cM over *SQSTM1* suggesting an estimated ninth degree of relatedness. Indeed, one pair within this trio shared an 11.442 cM IBD segment with relatedness estimated at the eighth degree. This relatedness suggests *SQSTM1* p.K238E is a low penetrance FALS mutation. A sixth degree relationship was identified between two *SOD1* p.I114T mutation carriers (online supplementary table 3; reported by Henden *et al*[35]). The remaining putative relationships identified were at the eighth degree or more distant, with IBD segments between 3 cM and 8 cM. Given that TRIBES relationship estimates greater than seventh degree are only 13% accurate,[33 35] and the shared ALS-implicated variants occur in the general population, it is less likely that these were related individuals. For the 41 SALS cases that carried a *C9orf72* repeat expansion and 10 SALS cases with an *ATXN2* intermediate expansion, no shared IBD segments >3 cM were identified.

An oligogenic and/or polygenic basis of ALS has been suggested by previous studies where multiple ALS gene variants have been identified in individual FALS cases[11–14] and SALS cases,[12 37 40] whereby variants may act together to cause ALS, or influence clinical manifestation. Our analysis showed 6.82% of SALS carried multiple ALS-implicated variants, further supporting an oligogenic/polygenic basis to ALS. In most reports of oligogenic ALS cases, one reported variant has been the pathogenic expansion of *C9orf72*.[11–14 37 40] Seventeen of the 42 (40.48%) putative oligogenic cases in our cohort had a *C9orf72* expansion. Over 40% of *C9orf72*-positive SALS cases in our cohort (n=41) had putative oligogenic disease. As described above, many of the ALS-implicated variants investigated here, and identified in a polygenic/oligogenic state, have limited evidence supporting their pathogenicity. Similarly, many variants identified in an oligogenic/polygenic state as part of other studies have limited evidence to support a pathogenic role. Functional studies will be necessary to determine whether particular variants act together to cause or modify the presentation of ALS.

It has been hypothesised that the development of ALS is a multistep process.[16] Oligogenic and polygenic factors may underlie steps in the development of ALS and explain the highly variable phenotypic presentation and course of disease, including that seen between cases who carry identical causal mutations.

As such, we sought to determine whether cases who harboured multiple ALS-implicated variants exhibited more severe clinical characteristics. We showed that patients with more than one ALS-implicated variant were significantly more likely to develop disease earlier in life than those with no known ALS-implicated variant. This is consistent with the concept that multiple 'hits' are required to trigger disease onset. It will be of interest for future studies to assess whether clinical features associate with specific mutant genes or with a specific variant or combination of variants. This will require large cohorts of SALS cases that carry mutations in the same ALS gene or identical ALS gene mutations.

Polygenic risk scores have been proposed as a tool to stratify patients with ALS, particularly SALS. Polygenic risk scores estimate an individual's risk of developing disease, based on the number of risk variants they carry and the relative disease risk each imposes. Both the phenotypic and genetic heterogeneity of ALS complicate the design of clinical trials, and in turn, the application of therapeutics to the spectrum of ALS cases.[2 4] Cohort stratification using polygenic risk scores has the potential to maximise the efficacy of therapeutic trials. While we were unable to calculate polygenic risk scores here due to insufficient sample size, our data, particularly our demonstration that patients carrying multiple ALS-implicated variants are more likely to have earlier disease onset, suggest that polygenic risk scores have potential to be useful in clinical settings and clinical trial design.

We used ExpansionHunter to directly parse NGS data for expansions in both *C9orf72* and *ATXN2*. The accuracy of ExpansionHunter for sizing *C9orf72* hexanucleotide expansions was originally reported at over 99.9%,[28] which was also reflected by our own validation data. Nevertheless, there remains the possibility of false negative cases. Southern blot analysis was not feasible for precise *C9orf72* expansion sizing, due to a lack of sufficient DNA from most cases, typical of historical sample cohorts.

In conclusion, we have explored the heterogeneity of SALS using a comprehensive survey of implicated ALS variants. Our data support a genetic predisposition to SALS with the presence of multiple ALS-implicated variants associated with earlier disease onset. Future studies are likely to continue to uncover novel genetic susceptibility and modifier variants that influence the development, presentation and course of SALS. It is hoped that the phenotypic and genetic characterisation of SALS will aid the design of biomarker studies and therapeutic trials.

Author affiliations
[1]Macquarie University Centre for Motor Neuron Disease Research, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South Wales, Australia
[2]Transformational Bioinformatics, Commonwealth Scientific and Industrial Research Organisation, Sydney, New South Wales, Australia
[3]Discipline of Pathology and Department of Neuropathology, The University of Sydney, Sydney, New South Wales, Australia
[4]Brain and Mind Centre, The University of Sydney, Sydney, New South Wales, Australia
[5]Department of Neuropathology, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia
[6]Institute of Clinical Neurosciences, Royal Prince Alfred Hospital, Sydney, New South Wales, Australia
[7]Department of Clinical Medicine, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South Wales, Australia

ORCID iDs
Emily P McCann http://orcid.org/0000-0002-8200-3525
Lyndal Henden http://orcid.org/0000-0001-5541-7637
Jennifer A Fifita http://orcid.org/0000-0002-0814-0118
Natalie Grima http://orcid.org/0000-0001-6434-7706
Denis C Bauer http://orcid.org/0000-0001-8033-9810
Sandrine Chan Moi Fat http://orcid.org/0000-0002-6304-0283
Natalie A Twine http://orcid.org/0000-0002-2051-8275
Roger Pamphlett http://orcid.org/0000-0003-3326-7273
Dominic B Rowe http://orcid.org/0000-0003-0912-2146
Kelly L Williams http://orcid.org/0000-0001-6319-9473
Ian P Blair http://orcid.org/0000-0003-3083-0075

REFERENCES

1 Leigh PN, Dodson A, Swash M, Brion JP, Anderton BH. Cytoskeletal abnormalities in motor neuron disease. An immunocytochemical study. *Brain* 1989;112:521–35.

2 Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. *Nat Rev Neurol* 2014;10:661–70.

3 Tiryaki E, Horak HA. Als and other motor neuron diseases. *Continuum* 2014;20:1185–207.

4 Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, Burrell JR, Zoing MC. Amyotrophic lateral sclerosis. *Lancet* 2011;377:942–55.

5 Ringholz GM, Appel SH, Bradshaw M, Cooke NA, Mosnik DM, Schulz PE. Prevalence and patterns of cognitive impairment in sporadic ALS. *Neurology* 2005;65:586–90.

6 Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* 2014;17:17–23.

7 McCann EP, Williams KL, Fifita JA, Tarr IS, O'Connor J, Rowe DB, Nicholson GA, Blair IP. The genotype-phenotype landscape of familial amyotrophic lateral sclerosis in Australia. *Clin Genet* 2017;92:259–66.

8 Al-Chalabi A, Fang F, Hanby MF, Leigh PN, Shaw CE, Ye W, Rijsdijk F. An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry* 2010;81:1324–6.

9 Wingo TS, Cutler DJ, Yarab N, Kelly CM, Glass JD. The heritability of amyotrophic lateral sclerosis in a clinically ascertained United States research registry. *PLoS One* 2011;6:e27985–10.

10 Ryan M, Heverin M, McLaughlin RL, Hardiman O. Lifetime risk and heritability of amyotrophic lateral sclerosis. *JAMA Neurol* 2019. doi:10.1001/jamaneurol.2019.2044

11 van Blitterswijk M, van Es MA, Hennekam EAM, Dooijes D, van Rheenen W, Medic J, Bourque PR, Schelhaas HJ, van der Kooi AJ, de Visser M, de Bakker PIW, Veldink JH, van den Berg LH. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet* 2012;21:3776–84.

12 Keogh MJ, Wei W, Aryaman J, Wilson I, Talbot K, Turner MR, McKenzie C-A, Troakes C, Attems J, Smith C, Al Sarraj S, Morris CM, Ansorge O, Pickering-Brown S, Jones N, Ironside JW, Chinnery PF. Oligogenic genetic variation of neurodegenerative disease genes in 980 postmortem human brains. *J Neurol Neurosurg Psychiatry* 2018;89:813–6.

13 Zou Z-Y, Zhou Z-R, Che C-H, Liu C-Y, He R-L, Huang H-P. Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2017;88:540–9.

14 Williams KL, Fifita JA, Vucic S, Durnall JC, Kiernan MC, Blair IP, Nicholson GA. Pathophysiological insights into ALS with C9orf72 expansions. *J Neurol Neurosurg Psychiatry* 2013;84:931–5.

15 van Rheenen W, Shatunov A, Dekker AM, McLaughlin RL, Diekstra FP, Pulit SL, van der Spek RAA, Võsa U, de Jong S, Robinson MR, Yang J, Fogh I, van Doormaal PT, Tazelaar GHP, Koppers M, Blokhuis AM, Sproviero W, Jones AR, Kenna KP, van Eijk KR, Harschnitz O, Schellevis RD, Brands WJ, Medic J, Menelaou A, Vajda A, Ticozzi N, Lin K, Rogelj B, Vrabec K, Ravnik-Glavač M, Koritnik B, Zidar J, Leonardis L, Grošelj LD, Millecamps S, Salachas F, Meininger V, de Carvalho M, Pinto S, Mora JS, Rojas-García R, Polak M, Chandran S, Colville S, Swingler R, Morrison KE, Shaw PJ, Hardy J, Orrell RW, Pittman A, Sidle K, Fratta P, Malaspina A, Topp S, Petri S, Abdulla S, Drepper C, Sendtner M, Meyer T, Ophoff RA, Staats KA, Wiedau-Pazos M, Lomen-Hoerth C, Van Deerlin VM, Trojanowski JQ, Elman L, McCluskey L, Basak AN, Tunca C, Hamzeiy H, Parman Y, Meitinger T, Lichtner P, Radivojkov-Blagojevic M, Andres CR, Maurel C, Bensimon G, Landwehrmeyer B, Brice A, Payan CAM, Saker-Delye S, Dürr A, Wood NW, Tittmann L, Lieb W, Franke A, Rietschel M, Cichon S, Nöthen MM, Amouyel P, Tzourio C, Dartigues J-F, Uitterlinden AG, Rivadeneira F, Estrada K, Hofman A, Curtis C, Blauw HM, van der Kooi AJ, de Visser M, Goris A, Weber M, Shaw CE, Smith BN, Pansarasa O, Cereda C, Del Bo R, Comi GP, D'Alfonso S, Bertolin C, Sorarù G, Mazzini L, Pensato V, Gellera C, Tiloca C, Ratti A, Calvo A, Moglia C, Brunetti M, Arcuti S, Capozzo R, Zecca C, Lunetta C, Penco S, Riva N, Padovani A, Filosto M, Muller B, Stuit RJ, Blair I, Zhang K, McCann EP, Fifita JA, Nicholson GA, Rowe DB, Pamphlett R, Kiernan MC, Grosskreutz J, Witte OW, Ringer T, Prell T, Stubendorff B, Kurth I, Hübner CA, Leigh PN, Casale F, Chio A, Beghi E, Pupillo E, Tortelli R, Logroscino G, Powell J, Ludolph AC, Weishaupt JH, Robberecht W, Van Damme P, Franke L, Pers TH, Brown RH, Glass JD, Landers JE, Hardiman O, Andersen PM, Corcia P, Vourc'h P, Silani V, Wray NR, Visscher PM, de Bakker PIW, van Es MA, Pasterkamp RJ, Lewis CM, Breen G, Al-Chalabi A, van den Berg LH, Veldink JH, PARALS Registry, SLALOM Group, SLAP Registry, FALS Sequencing Consortium, SLAGEN Consortium, NNIPPS Study Group. Genome-Wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat Genet* 2016;48:1043–8.

16 Al-Chalabi A, Calvo A, Chio A, Colville S, Ellis CM, Hardiman O, Heverin M, Howard RS, Huisman MHB, Keren N, Leigh PN, Mazzini L, Mora G, Orrell RW, Rooney J, Scott KM, Scotton WJ, Seelen M, Shaw CE, Sidle KS, Swingler R, Tsuda M, Veldink JH, Visser AE, van den Berg LH, Pearce N. Analysis of amyotrophic lateral sclerosis as a multistep process: a population-based modelling study. *Lancet Neurol* 2014;13:1108–13.

17 Chiò A, Mazzini L, D'Alfonso S, Corrado L, Canosa A, Moglia C, Manera U, Bersano E, Brunetti M, Barberis M, Veldink JH, van den Berg LH, Pearce N, Sproviero W, McLaughlin R, Vajda A, Hardiman O, Rooney J, Mora G, Calvo A, Al-Chalabi A. The multistep hypothesis of ALS revisited: the role of genetic mutations. *Neurology* 2018;91:e635–42.

18 Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Mot Neuron Disord* 2000;1:293–9.

19 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.

20 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.

21 Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013;43:1–33.

22 Chia R, Chiò A, Traynor BJ. Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. *Lancet Neurol* 2018;17:94–102.

23 Mackenzie IR, Nicholson AM, Sarkar M, Messing J, Purice MD, Pottier C, Annu K, Baker M, Perkerson RB, Kurti A, Matchett BJ, Mittag T, Temirov J, Hsiung G-YR, Krieger C, Murray ME, Kato M, Fryer JD, Petrucelli L, Zinman L, Weintraub S, Mesulam M, Keith J, Zivkovic SA, Hirsch-Reinshagen V, Roos RP, Zürcher S, Graff-Radford NR, Petersen RC, Caselli RJ, Wszolek ZK, Finger E, Lippa C, Lacomis D, Stewart H, Dickson DW, Kim HJ, Rogaeva E, Bigio E, Boylan KB, Taylor JP, Rademakers R. Tia1 mutations in amyotrophic lateral sclerosis and frontotemporal dementia promote phase separation and alter stress granule dynamics. *Neuron* 2017;95:808–16.

24 Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, Dominov JA, Kenna BJ, Nalls MA, Keagle P, Rivera AM, van Rheenen W, Murphy NA, van Vugt JJFA, Geiger JT, Van der Spek RA, Pliner HA, Smith BN, Marangi G, Topp SD, Abramzon Y, Gkazi AS, Eicher JD, Kenna A, Mora G, Calvo A, Mazzini L, Riva N, Mandrioli J, Caponnetto C, Battistini S, Volanti P, La Bella V, Conforti FL, Borghero G, Messina S, Simone IL, Trojsi F, Salvi F, Logullo FO, D'Alfonso S, Corrado L, Capasso M, Ferrucci L, Moreno CdeAM, Kamalakaran S, Goldstein DB, Gitler AD, Harris T, Myers RM, Phatnani H, Musunuri RL, Evani US, Abhyankar A, Zody MC, Kaye J, Finkbeiner S, Wyman SK, LeNail A, Lima L, Fraenkel E, Svendsen CN, Thompson LM, Van Eyk JE, Berry JD, Miller TM, Kolb SJ, Cudkowicz M, Baxi E, Benatar M, Taylor JP, Rampersaud E, Wu G, Wuu J, Lauria G, Verde F, Fogh I, Tiloca C, Comi GP, Sorarù G, Cereda C, Corcia P, Laaksovirta H, Myllykangas L, Jansson L, Valori M, Ealing J, Hamdalla H, Rollinson S, Pickering-Brown

S, Orrell RW, Sidle KC, Malaspina A, Hardy J, Singleton AB, Johnson JO, Arepalli S, Sapp PC, McKenna-Yasek D, Polak M, Asress S, Al-Sarraj S, King A, Troakes C, Vance C, de Belleroche J, Baas F, Ten Asbroek ALMA, Muñoz-Blanco JL, Hernandez DG, Ding J, Gibbs JR, Scholz SW, Floeter MK, Campbell RH, Landi F, Bowser R, Pulst SM, Ravits JM, MacGowan DJL, Kirby J, Pioro EP, Pamphlett R, Broach J, Gerhard G, Dunckley TL, Brady CB, Kowall NW, Troncoso JC, Le Ber I, Mouzat K, Lumbroso S, Heiman-Patterson TD, Kamel F, Van Den Bosch L, Baloh RH, Strom TM, Meitinger T, Shatunov A, Van Eijk KR, de Carvalho M, Kooyman M, Middelkoop B, Moisse M, McLaughlin RL, Van Es MA, Weber M, Boylan KB, Van Blitterswijk M, Rademakers R, Morrison KE, Basak AN, Mora JS, Drory VE, Shaw PJ, Turner MR, Talbot K, Hardiman O, Williams KL, Fifita JA, Nicholson GA, Blair IP, Rouleau GA, Esteban-Pérez J, García-Redondo A, Al-Chalabi A, Rogaeva E, Zinman L, Ostrow LW, Maragakis NJ, Rothstein JD, Simmons Z, Cooper-Knock J, Brice A, Goutman SA, Feldman EL, Gibson SB, Taroni F, Ratti A, Gellera C, Van Damme P, Robberecht W, Fratta P, Sabatelli M, Lunetta C, Ludolph AC, Andersen PM, Weishaupt JH, Camu W, Trojanowski JQ, Van Deerlin VM, Brown RH, van den Berg LH, Veldink JH, Harms MB, Glass JD, Stone DJ, Tienari P, Silani V, Chiò A, Shaw CE, Traynor BJ, Landers JE, ITALSGEN Consortium, Genomic Translation for ALS Care (GTAC) Consortium, ALS Sequencing Consortium, NYGC ALS Consortium, Answer ALS Foundation, Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATe) Consortium, SLAGEN Consortium, French ALS Consortium, Project MinE ALS Sequencing Consortium. Genome-Wide analyses identify KIF5A as a novel ALS gene. *Neuron* 2018;97:1268–83.

25 Benyamin B, He J, Zhao Q, Gratten J, Garton F, Leo PJ, Liu Z, Mangelsdorf M, Al-Chalabi A, Anderson L, Butler TJ, Chen L, Chen X-D, Cremin K, Deng H-W, Devine M, Edson J, Fifita JA, Furlong S, Han Y-Y, Harris J, Henders AK, Jeffree RL, Jin Z-B, Li Z, Li T, Li M, Lin Y, Liu X, Marshall M, McCann EP, Mowry BJ, Ngo ST, Pamphlett R, Ran S, Reutens DC, Rowe DB, Sachdev P, Shah S, Song S, Tan L-J, Tang L, van den Berg LH, van Rheenen W, Veldink JH, Wallace RH, Wheeler L, Williams KL, Wu J, Wu X, Yang J, Yue W, Zhang Z-H, Zhang D, Noakes PG, Blair IP, Henderson RD, McCombe PA, Visscher PM, Xu H, Bartlett PF, Brown MA, Wray NR, Fan D. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat Commun* 2017;8:611.

26 Abel O, Powell JF, Andersen PM, Al-Chalabi A. ALSoD: a user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. *Hum Mutat* 2012;33:1345–51.

27 Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Neale BM, Daly MJ, MacArthur DG, Consortium TGAD. Variation across 141, 456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 2019;531210.

28 Dolzhenko E, van Vugt JJFA, Shaw RJ, Bekritsky MA, van Blitterswijk M, Narzisi G, Ajay SS, Rajan V, Lajoie BR, Johnson NH, Kingsbury Z, Humphray SJ, Schellevis RD, Brands WJ, Baker M, Rademakers R, Kooyman M, Tazelaar GHP, van Es MA, McLaughlin R, Sproviero W, Shatunov A, Jones A, Al Khleifat A, Pittman A, Morgan S, Hardiman O, Al-Chalabi A, Shaw C, Smith B, Neo EJ, Morrison K, Shaw PJ, Reeves C, Winterkorn L, Wexler NS, Housman DE, Ng CW, Li AL, Taft RJ, van den Berg LH, Bentley DR, Veldink JH, Eberle MA, US–Venezuela Collaborative Research Group. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res* 2017;27:1895–903.

29 Dolzhenko E, Deshpande V, Schlesinger F, Krusche P, Petrovski R, Chen S, Emig-Agius D, Gross A, Narzisi G, Bowman B, Scheffler K, van Vugt JJFA, French C, Sanchis-Juan A,

Ibáñez K, Tucci A, Lajoie BR, Veldink JH, Raymond FL, Taft RJ, Bentley DR, Eberle MA. ExpansionHunter: a sequence-graph-based tool to analyze variation in short tandem repeat regions. *Bioinformatics* 2019;35:4754–6.

30 Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A, Abramzon Y, Remes AM, Kaganovich A, Scholz SW, Duckworth J, Ding J, Harmer DW, Hernandez DG, Johnson JO, Mok K, Ryten M, Trabzuni D, Guerreiro RJ, Orrell RW, Neal J, Murray A, Pearson J, Jansen IE, Sondervan D, Seelaar H, Blake D, Young K, Halliwell N, Callister JB, Toulson G, Richardson A, Gerhard A, Snowden J, Mann D, Neary D, Nalls MA, Peuralinna T, Jansson L, Isoviita V-M, Kaivorinne A-L, Hölttä-Vuori M, Ikonen E, Sulkava R, Benatar M, Wuu J, Chiò A, Restagno G, Borghero G, Sabatelli M, Heckerman D, Rogaeva E, Zinman L, Rothstein JD, Sendtner M, Drepper C, Eichler EE, Alkan C, Abdullaev Z, Pack SD, Dutra A, Pak E, Hardy J, Singleton A, Williams NM, Heutink P, Pickering-Brown S, Morris HR, Tienari PJ, Traynor BJ, Calvo A, Cammarosano S, Moglia C, Canosa A, Gallo S, Brunetti M, Ossola I, Mora G, Marinou K, Papetti L, Conte A, Luigetti M, Bella L V, Spataro R, Colletti T, Battistini S, Giannini F, Ricci C, Caponnetto C, Mancardi G, Mandich P, Salvi F, Bartolomei I, Mandrioli J, Sola P, Corbo M, Lunetta C, Penco S, Monsurro MR, Tedeschi G, Conforti FL, Volanti P, Floris G, Cannas A, Piras V, Murru MR, Marrosu MG, Pugliatti M, Ticca A, Simone I, Logroscino G, ITALSGEN Consortium. A hexanucleotide repeat expansion in C9orf72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011;72:257–68.

31 Van Damme P, Veldink JH, van Blitterswijk M, Corveleyn A, van Vught PWJ, Thijs V, Dubois B, Matthijs G, van den Berg LH, Robberecht W. Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. *Neurology* 2011;76:2066–72.

32 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.

33 Twine NA, Szul P, Henden L, McCann EP, Blair IP, Williams KL, Bauer DC. Tribes: a user-friendly pipeline for relatedness detection and disease gene discovery. *bioRxiv* 2019;686253.

34 Henden L, Lee S, Mueller I, Barry A, Bahlo M. Identity-By-Descent analyses for measuring population dynamics and selection in recombining pathogens. *PLoS Genet* 2018;14:e1007279.

35 Henden L, Twine NA, Szul P, McCann EP, Nicholson GA, Rowe DB, Kiernan MC, Bauer DC, Blair IP, Williams KL. Ibd analysis of Australian amyotrophic lateral sclerosis SOD1-mutation carriers identifies five founder events and links sporadic cases to existing ALS families. *bioRxiv* 2019;685925.

36 McCombe PA, Henderson RD. Effects of gender in amyotrophic lateral sclerosis. *Gend Med* 2010;7:557–70.

37 Black HA, Leighton DJ, Cleary EM, Rose E, Stephenson L, Colville S, Ross D, Warner J, Porteous M, Gorrie GH, Swingler R, Goldstein D, Harms MB, Connick P, Pal S, Aitman TJ, Chandran S. Genetic epidemiology of motor neuron disease-associated variants in the Scottish population. *Neurobiol Aging* 2017;51:178.e11–178.e20.

38 Kim H-J, Oh K-W, Kwon M-J, Oh S-I, Park J-S, Kim Y-E, Choi B-O, Lee S, Ki C-S, Kim SH, KW O, il OS, seok PJ, CS K. Identification of mutations in Korean patients with amyotrophic lateral sclerosis using multigene panel testing. *Neurobiol Aging* 2016;37:209.e9–209.e16.

39 Brenner D, Müller K, Wieland T, Weydt P, Böhm S, Lulé D, Hübers A, Neuwirth C, Weber M, Borck G, Wahlqvist M, Danzer KM, Volk AE, Meitinger T, Strom TM, Otto M, Kassubek J, Ludolph AC, Andersen PM, Weishaupt JH. Nek1 mutations in familial amyotrophic lateral sclerosis. *Brain* 2016;139:e28.

40 Morgan S, Shatunov A, Sproviero W, Jones AR, Shoai M, Hughes D, Al Khleifat A, Malaspina A, Morrison KE, Shaw PJ, Shaw CE, Sidle K, Orrell RW, Fratta P, Hardy J, Pittman A, Al-Chalabi A. A comprehensive analysis of rare genetic variation in amyotrophic lateral sclerosis in the UK. *Brain* 2017;140:1611–8.