Open Access

ORIGINAL ARTICLE

Whole exome sequencing and DNA methylation analysis in a clinical amyotrophic lateral sclerosis cohort

Fleur C. Garton[1,2] (iD), Beben Benyamin[1,2], Qiongyi Zhao[1], Zhijun Liu[1,2], Jacob Gratten[1,2], Anjali K. Henders[1,2], Zong-Hong Zhang[1], Janette Edson[1], Sarah Furlong[1], Sarah Morgan[3,4], Susan Heggie[5], Kathryn Thorpe[5], Casey Pfluger[5], Karen A. Mather[6], Perminder S. Sachdev[6,7], Allan F. McRae[1,2], Matthew R. Robinson[1,2], Sonia Shah[1], Peter M. Visscher[1,2,8], Marie Mangelsdorf[1], Robert D. Henderson[9], Naomi R. Wray[1,2,*] & Pamela A. McCombe[5,*]

[1]Queensland Brain Institute, The University of Queensland, St Lucia, Queensland 4072, Australia
[2]Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland 4072, Australia
[3]Reta Lila Weston Institute, UCL Institute of Neurology, London, UK
[4]Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
[5]UQ Centre for Clinical Research, The University of Queensland, Royal Brisbane & Women's Hospital, Brisbane 4029, Australia
[6]Centre for Healthy Brain Ageing, School of Psychiatry, Faculty of Medicine, The University of New South Wales, Sydney, New South Wales 2052, Australia
[7]Neuropsychiatric Institute, Prince of Wales Hospital, Randwick, New South Wales, Australia
[8]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4012, Australia
[9]Department of Neurology, Royal Brisbane & Women's Hospital, Brisbane 4029, Australia

Abstract

Background
Gene discovery has provided remarkable biological insights into amyotrophic lateral sclerosis (ALS). One challenge for clinical application of genetic testing is critical evaluation of the significance of reported variants.

Methods
We use whole exome sequencing (WES) to develop a clinically relevant approach to identify a subset of ALS patients harboring likely pathogenic mutations. In parallel, we assess if DNA methylation can be used to screen for pathogenicity of novel variants since a methylation signature has been shown to associate with the pathogenic *C9orf72* expansion, but has not been explored for other ALS mutations. Australian patients identified with ALS-relevant variants were cross-checked with population databases and case reports to critically assess whether they were "likely causal," "uncertain significance," or "unlikely causal."

Results
Published ALS variants were identified in >10% of patients; however, in only 3% of patients (4/120) could these be confidently considered pathogenic (in *SOD1* and *TARDBP*). We found no evidence for a differential DNA methylation signature in these mutation carriers.

Conclusions
The use of WES in a typical ALS clinic demonstrates a critical approach to variant assessment with the capability to combine cohorts to enhance the largely unknown genetic basis of ALS.

Introduction

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by the progressive loss of upper and lower motor neurons with diagnosis based on clinical and neurophysiological criteria. Typically, it presents as late-onset muscle weakness leading to paralysis and death within 3–5 years (Hardiman et al. 2011; Majounie et al. 2012), although there is considerable heterogeneity of motor and extramotor features (McCombe et al. 2016).

Amyotrophic lateral sclerosis likely arises from a combination of genetic susceptibility (Marangi and Traynor 2015; Renton et al. 2014) and environmental factors (Fang et al. 2009), but the contributing genetic variants can differ among individuals. About 10% of patients report a first degree relative with a diagnosis of ALS (familial ALS, fALS). Analyses of genetic data from large samples of fALS cases and nonaffected family members have led to the identification of mutations that segregate with disease, and these are collated in the ALS online Database (ALSoD) (Abel et al. 2012). Some recorded variants are pathogenic, or at least highly penetrant, and the latter include the *C9orf72* hexanucleotide (G_4C_2) repeat expansion (HRE) (DeJesus-Hernandez et al. 2011a; Renton et al. 2011a), the most common variant known to cause ALS (MIM#105550). Worldwide, the *C9orf72* HRE is estimated to account for ~38% fALS and ~6% of those presenting as sporadic (sALS, i.e., no or limited family history) with penetrance estimated as 50% by 58 years, and almost 100% by 80 years (Majounie et al. 2012). Three other genes (*SOD1*, *TARDBP*, and *FUS*) harbor pathogenic variants which together account for ~20% of fALS cases (Renton et al. 2014) (MIM#105400, #612069, #608030, respectively). In total, reported variants that fulfill adequate objective criteria for causation (Abel et al. 2013; MacArthur et al. 2014; Stenson et al. 2014) are in ~30 genes (Morgan et al. 2015). Most of these are autosomal dominant variants located in exonic regions of protein coding genes, however, many ALS-associated variants, even those in well-known genes, have not been sufficiently confirmed as pathogenic (Eisen et al. 2008). In fact, approximately 100 additional genes (658 known variants) are listed in ALSoD (Abel et al. 2012); the list includes rare variants requiring further evidence for proof of causality, as well as common variants that increase the risk of ALS. As discovery sample sizes increase, more variants are expected to be classified as pathogenic, while other variants in the database may be confirmed as false positives (benign). False positive reporting for variants segregating with disease is well recognized; for example, on average, each of the >60K individuals in the Exome Aggregation Consortium (ExAC) (Lek et al. 2016) study harbors 54 variants listed as causative of disease in the Mendelian disease database (OMIM). Moreover, larger sample sizes have allowed more accurate estimates of population frequencies of OMIM variants showing that 80% of them are too common to be compatible with disease pathogenicity.

With decreasing costs of genotyping technologies, implementation of routine genetic diagnostic testing in the ALS clinic is becoming more achievable and is desirable because earlier diagnosis could be achieved for a subset of individuals and a genetically informed diagnosis may allow better stratification of patients for therapy. For those presenting with sALS, ~11% harbor pathogenic variants (Renton et al. 2014). While changes in variant identification rely on the underlying etiologies of ALS, this proportion may increase as discovery samples increase to allow the list of confirmed variants to grow. Genetic testing for known causative mutations can be performed as single variant tests, or as a multiple variant test in a targeted capture, or through whole exome or genome sequencing (WES/WGS). The WES/WGS approach allows interrogation of the data for known mutations, as well as novel variants in causative genes, with periodic reanalysis of the data as new results become available. Effective filtering strategies with large control datasets are required to maximize true positives while reducing false positives. Other genetic technologies, such as DNA methylation studies, could contribute to predicting the pathogenicity of novel variants (Bonder et al. 2017). For example, carriers of a pathogenic HRE expansion in the *C9orf72* gene have been reported to carry a *C9orf72* methylation signature (Xi et al. 2013; He et al. 2015; Gijselinck et al. 2016). This phenomenon has not been explored for other ALS mutations.

Here, we set out to determine what conclusions could be drawn from a typical ALS clinic if all patients were put forward for generation of WES. We propose a filtering strategy to highlight relevant results to cover all modes of variant architecture and assess whether a methylation signature coexists with causal mutations.

Materials and Methods

Ethical compliance

The study was approved by the RBWH, University of Queensland, QIMR, and UNSW Research Ethics Committees.

Subjects

Australian ALS patients were recruited from the Royal Brisbane and Women's Hospital (RBWH) and were defined as definite or probable ALS according to the revised El Escorial criteria (Brooks et al. 2000) ($N = 120$). For the methylation capture, patients were recruited as above with healthy controls sourced from Brisbane and Sydney across three different sites: RBWH, Queensland Institute of Medical Research (QIMR), and University of New South Wales (UNSW) (see Supporting Information).

Genetic data

Whole exome sequencing and genome-wide methylation

Briefly, patient DNA was extracted from fresh whole blood using manual extraction protocols. In controls obtained from UNSW, the majority (90%) of DNA was extracted from frozen whole blood or lymphocytes using a kit (Qiagen Autopure, Hilden, Germany). Sequencing libraries were prepared using the KAPA HYPER Prep kit (KAPA Biosystems, Wilmington, MA, USA) and followed by targeted exome capture using the Nimblegen SeqCap EZ Exome v3 (64 Mb) kit (Roche, Madison, WI, USA), both according to the manufacturer's instructions. Patient samples were sequenced in a 126-bp paired end mode using the Illumina HiSeq 2000 platform. Processing and sequence alignment were by standard protocols (GATK pipeline) (McKenna et al. 2010), and variants were annotated using the ANNOVAR software tool (version 2015 Jun17) (Wang et al. 2010). Following variant call and annotation, standard quality control (QC) steps were implemented, that is, common SNPs to remove individuals identified as being discordant for sex, excessive heterozygosity, or relatedness. Genome-wide methylation data were generated using the Illumina HumanMethylation450 array, which contains probes covering 99% of reference sequences (RefSeq) genes and 96% of CpG islands. A standard QC pipeline was applied to remove probes bound to multiple locations and duplicate individuals (Supporting Information).

WES variant filtering and statistical analysis

The pathogenicity of single-nucleotide variants was evaluated based on a number of filtering steps (Supporting Information). Briefly, known variants and relevant genes were compiled from two databases, the human gene mutation database (HGMD) (Stenson et al. 2014) and the amyotrophic lateral sclerosis online database (ALSoD)
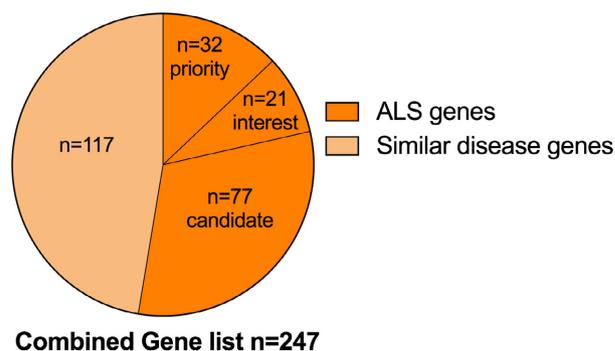
Figure 1. Combined gene list ($n = 247$) used for variant filtering. Genes associated with ALS were manually curated from ALSoD and HGMD and divided into independent subgroups (priority, interest, candidate). Also considered were genes with a high level of evidence for pathogenicity in diseases showing evidence of clinical overlap with ALS (frontotemporal dementia, spinal muscular atrophy, spinal bulbar muscular atrophy, distal hereditary motor neuropathy, Charcot–Marie–Tooth disease, hereditary spastic paraplegia, hereditary ataxia, and distal myopathy). Gene lists are found in Table S1.

(Abel et al. 2012). This consisted of three reference lists to examine patient exomes for known variants or novel variants in relevant genes (Fig. 1; Table S1), these lists were:

1 Variants associated with ALS ($n = 1158$ variants)
2 Variants in known ALS genes characterized based on literature review as priority ($n = 32$ genes), interest ($n = 21$ genes), and candidate ($n = 77$ genes). Pathogenic insertion and deletions (indels) were analyzed in a subset of genes ($n = 21$ genes, Table S1) previously published to contain indels or loss-of-function mutations known to cause ALS symptoms.
3 Variants in related disease genes, that is, genes with evidence of causation for diseases with genetic/symptomatic links to ALS, including frontotemporal dementia, spinal muscular atrophy, spinal bulbar muscular atrophy, distal hereditary motor neuropathy, Charcot–Marie–Tooth disease, hereditary spastic paraplegia, hereditary ataxia, and distal myopathy (HGMD, $n = 117$ genes).

These lists were hierarchical, that is, variants filtered in list (1) were not included in list (2).

Based on the latest recommendations of ExAC, novel variants were filtered out if their minor allele frequency (MAF) in any of the 15 available reference populations (Supporting Information) was above the specified threshold (Lek et al. 2016). This is referred to as the "population maximum" or "popmax" approach, whereby a variant is considered unlikely to be pathogenic if it has achieved the threshold frequency of ≥ 0.00005 for autosomal dominant and ≥ 0.01 for recessive/compound

heterozygous candidate variants, respectively (Cirulli et al. 2015), in any ancestry-specific population. This applies a more stringent threshold than using an average frequency across all populations to reduce the number of false positives. The final list of variants that passed the popmax filter included splice site mutations and nonsynonymous variants predicted to be deleterious by a combined evaluation of 18 current deleteriousness scoring methods as implemented in MetaLR and MetaSVM (Dong et al. 2015). Variants identified in our patients were then cross checked with the literature and online databases for case reports to assess whether they were "likely causal," "uncertain significance," or in the case of a known variant with high minor allele frequency "unlikely causal."

We were underpowered to perform either burden testing or gene-set analyses; however, exome data will be uploaded to ALSdb (V2, New York City, New York [URL: http://alsdb.org]) to contribute to future larger studies. Likely causal pathogenic variants and anonymous patient data will be reported in the ALSoD (Abel et al. 2012) patient database.

Detection of DNA methylation outliers

From previous analyses and published results, 40–60% of *C9orf72* HRE carriers have been shown to be extreme DNA methylation outliers in blood (Xi et al. 2013; He et al. 2015; Gijselinck et al. 2016). Since genome-wide association studies have identified single- nucleotide polymorphisms (SNPs) associated with variation in DNA methylation between individuals (Bonder et al. 2017), our hypothesis was that carriers of pathogenic variants in known ALS genes may exhibit hypo- or hypermethylation in the same gene, relative to noncarriers. For an initial screening, carriers were considered outliers if they lay $1.5\times$ outside the interquartile range of noncarriers. Analysis particularly focused on probes closest to the transcription start site (TSS) of the single-nucleotide variant gene. More in-depth analysis was not justified based on the results of this initial screening.

Results

A summary of clinical data for the 120 unrelated patients is given in Table 1. Briefly, the cohort included 75 males and 45 females, median age of onset was 61 years (SD = 10.1) and survival was 2.9 years (SD = 2.4) ($n = 80$). For WES, ~741.48 Gbp of sequence data were generated for these individuals, with a mean coverage of $53.9\times$ per individual. For DNA methylation analysis, 118 cases and 111 unrelated controls and 432,216 probes were used in analysis.

Table 1. Characteristics of ALS participants.

Number of ALS cases	120
Number of males	75 (62.5%)
Population in clinic capture cohort[1]	3.68 million
European ancestry	113 (93.0%)
Number reported to have first degree relatives with ALS/FTD (i.e., parent/sibling)	10 (8.3%)
Number reported to have second/third degree relative with ALS/FTD (i.e., grandparent/avuncular/ cousin) and no first degree relatives.	2 (1.6%)
Number reported a family history of related disease (dementia/MS/parkinsonism)	5 (4.1%)
Family history unknown	14 (11.5%)
Number of *C9orf72* HRE[2] positive	9 (7.5%)
Number of slow progressing >7 years	18 (15.0%)
Median (range) age at onset (years)	61 (SD ±10.1)
Median years survival (at death or censoring, $n = 80$)	2.9 (SD ±2.4)
Onset location[3]	
Number bulbar onset	30/118 (25%)
Number upper limb onset	33/118 (28%)
Number lower limb onset	53/118 (45%)
Number bulbar + Lower limb onset	1/118 (1%)
Number bulbar + Upper limb onset	1/118 (1%)

SD, standard deviation.

[1]Population estimate based on Northern Rivers NSW shires and Southeast QLD shires.

[2]See Supporting Information for details.

[3]Two individuals did not have a specified onset location. Family history reflects subject self-report of a family member known to have had ALS, FTD, or a related disease. Those that did not know their family history were reported as unknown.

Filtering of WES variants against reference list 1: variants reported to be associated with ALS

From 120 ALS patients, 17 variants were listed in either ALSoD or HGMD databases.

Four subjects (4/120 = 3%) were found to harbor variants classified as likely pathogenic, including two variants in *SOD1*, p.Gly94Val and p.Ile114Thr, and two variants in the *TARDBP*, p.Met337Val and p.Ala382Thr (Tables 2 and S2). Clinical data for these patients and comparison to available clinical notes from previous reports are listed in the Supporting Information. Although the remaining 13 variants were listed in the databases, each of these had a popmax MAF of ≥0.002, which is incompatible with current estimates of ALS lifetime risk and therefore are unlikely to be causal (Table S3). Half of these variants were reported by Daoud et al. (2011) based on a candidate gene analysis that is likely to be underpowered for identification of causal rare variants (380 cases/controls). A *NEK1* variant identified in our cohort (p.Arg261His) increases the risk

Table 2. Previously reported pathogenic variants found in Brisbane ALS patients.

Gene	Exon, nucleotide, amino acid change	Positive number/ negative number	Popmax	SNP	References
SOD1	exon4:c.281G>T:p.Gly94Val NM_000454.4[1]	1/119	NA		Brown et al. (2012)
SOD1	exon4:c.341T>C:p.Ile114Thr NM_000454.4[1]	1/119	NA	rs121912441	Brown et al. (2012), Rosen et al. (1993)
TARDBP	exon6:c.1009A>G:p.Met337Val NM_007375.3	1/119	NA	rs80356730	Tamaoka et al. (2010)
TARDBP	exon6:c.1144G>A:p.Ala382Thr NM_007375.3[1]	1/119	NA	rs367543041	Borghero et al. (2014), Quadri et al. (2011)

Positive number/negative number = number of cases identified with the alternate allele compared to those cases tested with the reference allele. All changes are single allele changes (heterozygous). Popmax = population maximum; frequency of the alternate allele in available control populations. A SNP rs number means that the SNP has been previously reported, but NA for popmax frequency means that the SNP is not listed in the databases and hence is assumed to be very rare. NCBI reference sequence provided.

[1]Variant independently confirmed by an Australian National Association of Testing Authorities (NATA) accredited laboratory.

of ALS (OR=2.4, 95% CI 1.6–3.7) (Kenna et al. 2016) but is not fully penetrant. No autosomal recessive or compound heterozygous genotypes of the listed ALS variants were identified.

Filtering of WES variants against reference list 2: variants in ALS-associated genes

For this filtering step we excluded patients who were positive for the likely pathogenic SOD1 (n = 2), TARDBP (n = 2), or C9orf72 (n = 9) so that 107 patients were examined for novel variants identified in known ALS genes. From the initial variant output of 2490, ~89% variants were too common to cause ALS with 2% being rare exonic changes. Variants that were predicted to be damaging included six in genes we labeled "priority", three in "interest" genes, and seven in "candidate" genes (Tables 3 and S4).

Indel variants examined in a subset of 21 genes known to cause ALS or ALS symptoms including a frameshift deletion (not reported before) caused by a single base deletion in exon 8 (p.277) of the TBK1 gene (Table 3). Using WES data to identify indels can produce high numbers of false positives (Fang et al. 2014) and Sanger sequencing is required for confirmation. If the deletion variant is confirmed, current evidence on the TBK1 deletion variant profile across multiethnic ALS and/or FTD cohorts suggests this could be pathogenic (Cirulli et al. 2015; Oakes et al. 2017).

For recessive variants, the MAF filter (0.01) removed 60% of the variants. A further 25% were excluded as they were not exonic/missense or splice site variants. No subsequent recessive variants were identified in any patients.

Table 3. Novel damaging variants in ALS genes found in Brisbane ALS patients.

Gene	Exon, nucleotide, amino acid changes	Positive number/ negative number	Popmax	SNP
DCTN1	exon19:c.2244C>G:p.Asp748Glu NM_004082.4	1/118	0.00003	–
DCTN1	exon14:c.1486G>C:p.Val496Leu NM_004082.4	1/84	0.00003	–
FUS	exon3:c.101G>A:p.Gly34Glu NM_004960.3	1/119	0	–
NEFH	exon4:c.2176G>C:p.Val726Leu NM_021076.3	1/119	0	–
SPG11	exon26:c.4557T>A:p.Asp1519Glu NM_025137.3	1/119	0.00001	–
TBK1	exon8:c.830delT:p.Leu277fs NM_013254.1	1/119	0	–

Positive number/negative number = number of cases identified with the alternate allele compared to those cases tested with the reference allele. All changes are single allele changes (heterozygous). Popmax = population maximum; frequency of the alternate allele in available control populations. NCBI reference sequence provided. None of the variants had a reference SNP cluster ID (SNP).

Filtering of WES variants against reference list 3: variants in related-disease genes

A total of 2290 variants were identified in the 117 genes harboring causal variants for diseases identified to have genetic/symptomatic cross-over with ALS (Renton et al. 2014). Ninety-one percent of variants were removed using a MAF of 0.00005, with a further 7% and 1% subsequently removed for being outside exons or synonymous

SNPs, respectively. Twenty-five autosomal dominant changes were identified in ALS cases with 18 of these predicted to be damaging. One was classified as "likely causal" (Supporting Information) with previous reports indicating pathogenicity for late onset distal myopathy (Fiorillo et al. 2016) and the other 17 were of "uncertain significance" (Table S5). The relevance of these results will only become apparent when similar analyses are undertaken with larger samples. No individuals were identified as being homozygous for any variant.

Identification of compound heterozygous variants in all genes

Three individuals were investigated for compound heterozygosity, because they have two rare variants in the same gene. The presence of two deleterious variants that occur one on each chromosome within a gene locus could result in complete gene dysfunction, which is known as compound heterozygosity. Alternatively, two deleterious variants may exist on the same chromosome to leave one functional allele (heterozygous mutations). For one case, the aligned sequencing reads revealed that the two variants (in *PNF1*) occurred always (73x) sequentially (i.e., on the same chromosome) and thus this individual was not compound heterozygous. The compound allele has been reported in both cases and controls, and considered to be risk variants rather than highly penetrant variants (Wu et al. 2012). One individual had two variants in the *TPP1* gene, namely a splice and a missense in exon 10. The same splice but different missense (exon 11) combination has been reported to cause spinocerebellar ataxia in a Dutch family (Sun et al. 2013). Another individual had two missense variants in the *AARS* gene occurring in the aminoacylation and tRNA recognition domain. Variants in this gene have been associated with CMT (Bansagi et al. 2015) and infantile encephalopathy (Simons et al. 2015) (Table S3).

Causative variant and methylation signature analysis

Here, we hypothesized that carriers of pathogenic variants (i.e., *C9orf72*, *SOD1*, and *TARDBP*) may exhibit hypo- or hypermethylation in the same gene, relative to noncarriers. Carriers were considered outliers if they lay outside of 1.5× the interquartile range of the noncarriers. We first investigated the DNA methylation signature for nine *C9orf72* HRE-positive ALS cases (Supporting Information). Two were identified as being methylation outliers in the two probes in the CpG island upstream of the transcriptional start site and the probe nearest to the HRE (Fig. S1). Having accurate information on the length of the expansions in

individuals would allow investigation of the relation of the repeat count and to methylation outlier status.

Given that the two *SOD1* and two *TARDBP* variants identified in ALS cases were considered to have the greatest evidence for ALS causality, the methylation signature (probes closest to the TSS or the variant) was assessed in each gene. Neither gene presented compelling evidence that these rare single-nucleotide variants were associated with differential DNA methylation in blood (Figs S1 and S2). For *SOD1*, probe cg03524836 located downstream of the variant demonstrated hypermethylation in one of two carriers, whereas for *TARDBP*, there was no evidence for altered methylation for either the closest TSS probe nor the closest variant probe in the *TARDBP* carrier.

Discussion

Testing for known variants affecting ALS could assist with ALS diagnosis in a subset of those presenting to an ALS clinic. We identified four patients with previously reported *SOD1* and *TARDBP* variants, and for each the time from onset to diagnosis was 1 year. In the future, early diagnosis could have impact on clinical practice and stratification for treatment. For these individuals, variants identified have been called "likely pathogenic" using an available evidence framework (Richards et al. 2015; Li and Wang 2017). In a clinical setting, we believe reporting of such variants would be considered useful as long as presented to the patients and families with the objective interpretation through genetic counseling. Grouping of patients by pathogenic mutation could elucidate associated clinical phenotypes. To use WES as a method to screen for genetic variants in a clinic population, a critical task involves identifying relevant genes/variants and whether they are pathogenic (MacArthur et al. 2014). Here, we present a tiered strategy to identify both known variants and novel variants in ALS patients (Fig. 1) followed by an evidence-based evaluation on whether genetic variants are likely causal or of uncertain significance (Fig. 2).

Sequencing whole exomes of our ALS case cohort has identified an additional 3% to harbor causal genetic variants – taking the overall genetic diagnosis of the cohort to 10% (accounting for the previously identified *C9orf72* HRE screening). A number of other candidate variants of uncertain significance were identified (20% ALS cases), but this list is likely to include false positives. Although there has been increasing interest in the possible role of oligogenic inheritance in ALS (i.e., more than one causative variants across more than one gene) (van Blitterswijk et al. 2012), we did not find evidence to support this in our cohort as no individual carried more than one "likely pathogenic" variant. Two individuals each carried two "uncertain significance" variants (Table S5); however, this
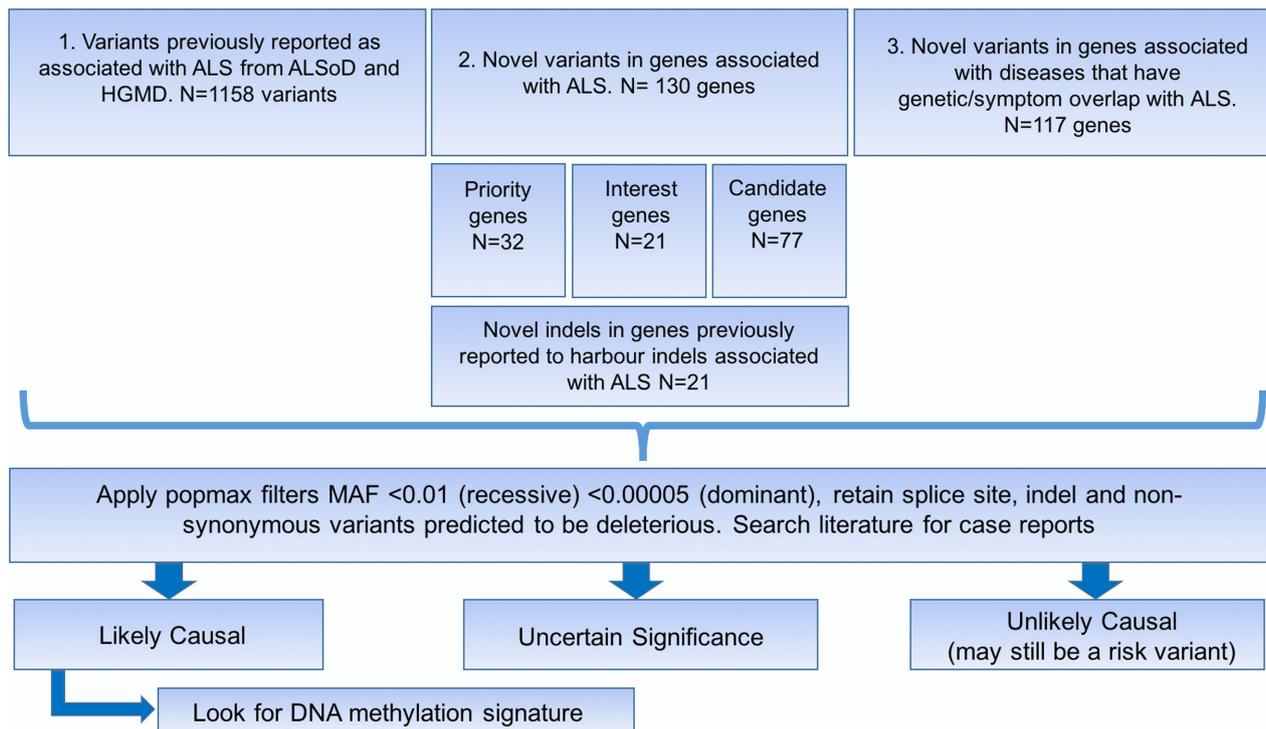
Figure 2. Flowchart demonstrating variant filtering methods and link with methylation data. After quality control (QC), whole exome sequences were analyzed for reported variants, novel variants in known genes, or novel variants in similar disease genes. ALS genes were divided into priority, interest, and candidate genes. This included a set of genes analyzed for the presence of insertion/deletion variants. Variants were subsequently filtered for rare, damaging variants with minor allele frequencies below the defined threshold before being prioritized based on prediction or literature. DNA methylation was collected independently and analyzed based on causal variant results. MAF, minor allele frequency.

list is likely to include false positives and thus no further conclusions can be drawn without further evidence.

The WES diagnostic rate was higher when considering only fALS cases ($n = 4/12$, Supporting Information) which is consistent with other studies (McCann et al. 2017). This relatively high diagnostic rate (30%), however, does not represent a typical clinic population as familial cases consist of a small proportion presenting cases. As WES case–control sample sizes increase, better separation of true and false positives and a clearer picture of the genetic architecture of ALS (in terms of population frequencies and effect sizes of variants) will be achieved. For a late-onset disorder such as ALS, it is relevant to note that presence of a variant in control databases does not preclude pathogenicity, however, we find, as previously reported, that many variants reported in databases as "pathogenic" occur at a population frequency too high to be consistent with the reported lifetime risk (Richards et al. 2015; Lek et al. 2016). Here, we used a population maximum frequency consistent with current guidelines (Lek et al. 2016) to classify 10% of variants as being "unlikely causal" (Fig. 3) (this does not preclude them from being risk variants).

We found variants that have been previously reported and show high evidence of causality in four patients (~3%), while we found an additional 5% and 15% of ALS patients harbor variants for which evidence is less strong (uncertain significance in ALS genes and related disease genes, respectively). In other genetic screening studies in clinic populations, variant filtering has been more lenient, typically using a MAF of 0.01 or 0.0001 (Table S6), resulting in the reporting of a greater number of identified variants, which could include a greater number of false positives. Any human population sample, whether an ALS cohort (Cirulli et al. 2015; Ozoguz et al. 2015) or other disease cohort (Lek et al. 2016), carry many rare genetic variants. If not efficiently examined, this heterogeneity can increase the difficulty and cost of genetic diagnosis, which remains a key barrier for implementation in the clinical setting (Roggenbuck et al. 2016). The use of single variant tests in clinical diagnostic laboratories is limited for ALS given the breadth of disease-causing variants that have been identified. Sequencing a common set of well-defined variants or genes, that is, targeted capture of the *SOD1* gene and specific exons of other genes (exon 6 of *TARDBP*, exons 14 and 15 of
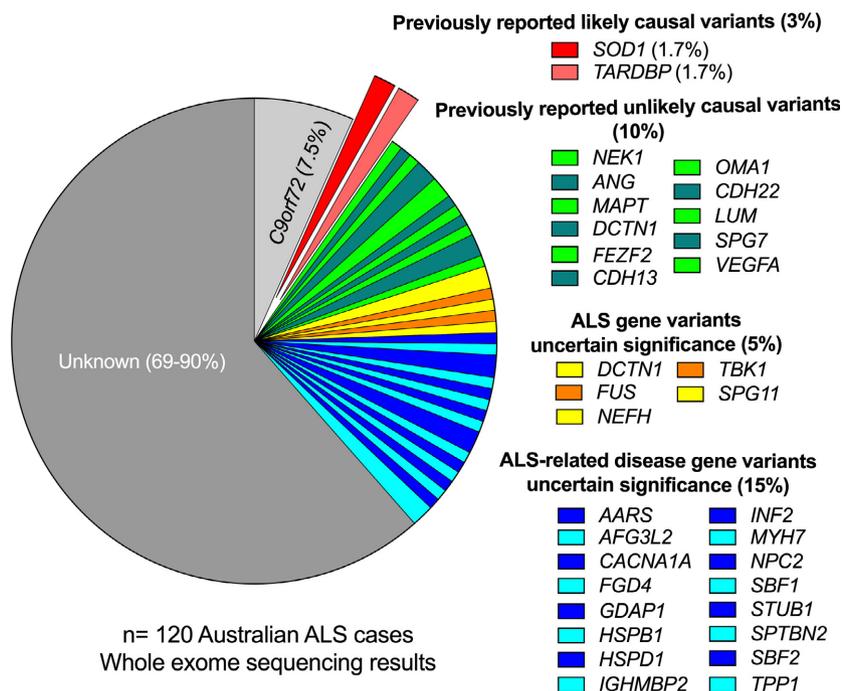
Figure 3. Summary of the whole exome sequencing results for Australian ALS case cohort. Variants identified were found in $n = 1$ or 2 individuals. Three percent of the cohort had a likely causal variant (exploded slices). Ten percent were positive for a previously reported ALS variant, but these were unlikely to be pathogenic because of their high population frequency. Novel variants (MAF <0.00005 dominant, <0.01 recessive), predicted to be damaging by either MetaSVM or MetaLR in ALS genes, were found in 5% of the cohort, with variants in ALS-related disease genes being found in 15% of the cohort. Most cases did not have a known variant or novel variant in ALS genes or ALS-related disease genes.

FUS), is a strategy currently being trialed (Table S6). In ALS, this portfolio of recurring mutation sites is expected to explain <5% of the cohort. With a rapidly expanding genetic profile of ALS-associated variants (Bettencourt and Houlden 2015) the WES approach is a long-term cost-effective genetic testing technique and first choice of a number of studies (Kenna et al. 2013; Agarwal et al. 2016). Despite requirements for confirmation testing (in a manner that satisfies the clinic's local jurisdiction that oversee these processes), with sufficient coverage, the majority of known ALS variants can be identified using this methodology.

A notable exception to variant detection with WES is the *C9orf72* HRE, which represents a significant proportion of ALS patients in familial (38%) and sporadic cases (6%) (Majounie et al. 2012). Carriers can be identified using Southern blotting (DeJesus-Hernandez et al. 2011b) or using a repeat-primed PCR method (Renton et al. 2011b). More recently, a signature of extreme DNA methylation has been detected in carriers compared to controls. While the lack of sensitivity (i.e., does not identify all carriers) means that it cannot replace traditional detection methods, it does have a high specificity (i.e., if the individual has altered methylation they are likely

to be a carrier) (Xi et al. 2013; He et al. 2015; Gijselinck et al. 2016). Here, we investigated if an altered methylation profile exists for other ALS variants. DNA methylation-based approaches to disease diagnosis in ALS are yet to be fully explored or understood. Since SNPs have been associated with variation in DNA methylation levels (mQTL) (Bonder et al. 2017), we were interested to investigate whether altered methylation may also be an effect of a pathogenic protein/disease variant. The preliminary investigations reported here on *SOD1* and *TARDBP* suggest that DNA methylation signatures from blood are unlikely to help in assessing functionality or penetrance of these rare single-nucleotide variants. Larger studies carried out in populations ideally with high frequency, likely causal variants, such as the Sardinian population for the *TARDBP* p.Ala382Thr (Orrù et al. 2012; Borghero et al. 2014), are needed to draw strong conclusions about the ability of DNA methylation outliers to help localize causal mutations and/or explain their penetrance. This includes how DNA sequence variants might alter methylation and gene expression outside the immediate location of the particular variant.

Research into the genetic basis of ALS has been rapidly advancing, identifying not only highly penetrant rare

variants (our focus here, and the proportion of occurrences of ALS cases explained by such variants will become clearer as sequencing studies grow) but also an additive background of more common risk variants. Substantial variation exists in attitude and practices related to genetic testing in the clinic (Arthur et al. 2017; Vajda et al. 2017) and a consistent approach to the collection and analysis of genetic data is needed in order to build a solid framework to underpin discovery and translation to clinical practice. Understanding the variability in genetic architecture across individuals with continued resource investment (detection of both rare and common associated variants through analysis of large samples) would contribute data to the possible development of genetic score predictors to aid in early diagnosis, stratify heterogeneity in clinical trials, and help design precise treatment pathways. This should encompass both familial (10%) and sporadic cases (90%) as it is well recognized that both types of presentations have a genetic component contributing to and/or altering their ALS diagnosis and symptom presentation.

Conclusions

We present an approach to the possible clinical use of whole exome sequencing in an Australian ALS case cohort of both familial and sporadic ALS cases. Applying this technology to a clinical cohort, we found that for 3% (four cases) it provided genetic information that could contribute to diagnosis of ALS. The *C9orf72* HRE remains the most common highly penetrant ALS-associated variant (~7% cases in European ancestry), and while this cannot currently be identified using WES, new evidence suggests this variant can be identified by whole genome sequencing (Dolzhenko et al. 2016). The application of next-generation sequencing to patient samples at the clinic can provide an efficient disease diagnosis for a small subset of individuals that has implications for the patient's family and in the future may guide treatment decisions tailored to the different causative variants. However, careful genetic counseling will be needed to convey the interpretation of putative pathogenic variants in those presenting without clear family history. As sample sizes for discovery of ALS-associated loci increase, particularly with large consortium approaches such as Project MinE, we believe improved diagnostic rates and genetic risk predictors will be achieved, and a clearer picture will emerge for their overall utility in the clinical setting. Contributions of DNA methylation approaches to disease diagnosis have not been yet sufficiently investigated, but these preliminary investigations on *SOD1* and *TARDBP* did not suggest that DNA methylation signatures from blood are helpful in assessing functionality of

these rare single-nucleotide variants. Our findings present a strategic method for pathogenic variant identification, with rapid results that can contribute to further investigation on risk susceptibility and causal variants contributing to ALS.

Conflict of Interest

None declared.

References

Abel, O., J. F. Powell, P. M. Andersen, et al. 2012. ALSoD: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. Hum. Mutat. 33:1345–1351.

Abel, O., J. F. Powell, P. M. Andersen, et al. 2013. Credibility analysis of putative disease-causing genes using bioinformatics. PLoS One 8:e64899.

Agarwal, S., L. Potocki, T. R. Collier, S. L. Woodbury, A. M. Adesina, J. Jones, et al. 2016. Utility of whole exome sequencing in evaluation of juvenile motor neuron disease. Muscle Nerve 53:648–652.

Arthur, K. C., C. Doyle, A. Chio, et al. 2017. Use of genetic testing in amyotrophic lateral sclerosis by neurologists. JAMA Neurol. 74:125–126.

Bansagi, B., T. Antoniadi, S. Burton-Jones, S. M. Murphy, J. Mchugh, M. Alexander, et al. 2015. Genotype/phenotype correlations in AARS-related neuropathy in a cohort of patients from the United Kingdom and Ireland. J. Neurol. 262:1899–1908.

Bettencourt, C., and H. Houlden. 2015. Exome sequencing uncovers hidden pathways in familial and sporadic ALS. Nat. Neurosci. 18:611–613.

van Blitterswijk, M., M. A. Van Es, E. A. Hennekam, D. Dooijes, W. Van Rheenen, J. Medic, et al. 2012. Evidence for an oligogenic basis of amyotrophic lateral sclerosis. Hum. Mol. Genet. 21:3776–3784.

Bonder, M. J., R. Luijk, D. V. Zhernakova, M. Moed, P. Deelen, M. Vermaat, et al. 2017. Disease variants alter transcription factor levels and methylation of their binding sites. Nat. Genet. 49:131–138.

Borghero, G., M. Pugliatti, F. Marrosu, M. G. Marrosu, M. R. Murru, G. Floris, et al. 2014. Genetic architecture of ALS in Sardinia. Neurobiol. Aging 35:2882.e2887–2882.e2812.

Brooks, B. R., R. G. Miller, M. Swash, et al. 2000. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyot Lat Sclerosis mot Neuron Dis 1: 293–299.

Brown, J. A., J. Min, J. F. Staropoli, E. Collin, S. Bi, X. Feng, et al. 2012. SOD1, ANG, TARDBP and FUS mutations in amyotrophic lateral sclerosis: a United States clinical testing lab experience. Amyot Lat Scleros 13:217–222.

Cirulli, E. T., B. N. Lasseigne, S. Petrovski, P. C. Sapp, P. A. Dion, C. S. Leblond, et al. 2015. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science (New York, NY) 347:1436–1441.

Daoud, H., P. N. Valdmanis, F. Gros-Louis, V. Belzil, D. Spiegelman, E. Henrion, et al. 2011. Resequencing of 29 candidate genes in patients with familial and sporadic amyotrophic lateral sclerosis. Arch. Neurol. 68:587–593.

Dejesus-Hernandez, M., I. R. Mackenzie, B. F. Boeve, A. L. Boxer, M. Baker, N. J. Rutherford, et al. 2011a. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome
9p-linked FTD and ALS. Neuron 72:245–256.

Dolzhenko, E., J. J. F. A. Van Vugt, R. J. Shaw, M. A. Bekritsky, M. Van Blitterswijk, Z. Kingsbury, et al. 2016. Detection of long repeat expansions from PCR-free whole-genome sequence data. bioRxiv 093831; doi: https://doi.org/10.1101/093831.

Dong, C., P. Wei, X. Jian, R. Gibbs, E. Boerwinkle, K. Wang, et al. 2015. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum. Mol. Genet. 24:2125–2137.

Eisen, A., M. M. Mezei, H. G. Stewart, et al. 2008. SOD1 gene mutations in ALS patients from British Columbia, Canada: clinical features, neurophysiology and ethical issues in management. Amyot. Lat. Scleros. 9:108–119.

Fang, F., F. Kamel, P. Lichtenstein, R. Bellocco, P. Sparen, D. P. Sandler, et al. 2009. Familial aggregation of amyotrophic lateral sclerosis. Ann. Neurol. 66:94–99.

Fang, H., Y. Wu, G. Narzisi, J. A. O'rawe, L. T. Barron, J. Rosenbaum, et al. 2014. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med. 6:89.

Fiorillo, C., G. Astrea, M. Savarese, D. Cassandrini, G. Brisca, F. Trucco, et al. 2016. MYH7-related myopathies: clinical, histopathological and imaging findings in a cohort of Italian patients. Orphanet. J. Rare Dis. 11:91.

Gijselinck, I., S. Van Mossevelde, J. Van Der Zee, A. Sieben, S. Engelborghs, J. De Bleecker, et al. 2016. The C9orf72 repeat

size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. Mol. Psychiatry 21:1112–1124.

Hardiman, O., L. H. van den Berg, and M. C. Kiernan. 2011. Clinical diagnosis and management of amyotrophic lateral sclerosis. Nat. Rev. Neurol. 7:639–649.

He, J., L. Tang, B. Benyamin, S. Shah, G. Hemani, R. Liu, et al. 2015. C9orf72 hexanucleotide repeat expansions in Chinese sporadic amyotrophic lateral sclerosis. Neurobiol. Aging 36:2660.e2661-2660.e2668.

Kenna, K. P., R. L. Mclaughlin, S. Byrne, M. Elamin, M. Heverin, E. M. Kenny, et al. 2013. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J. Med. Genet. 50:776–783.

Kenna, K. P., P. T. Van Doormaal, A. M. Dekker, N. Ticozzi, B. J. Kenna, F. P. Diekstra, et al. 2016. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat. Genet. 48:1037–1042.

Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285–291.

Li, Q., and K. Wang. 2017. InterVar: Clinical interpretation of genetic variants by the 2015 ACMG-AMP guidelines. Am. J. Hum. Genet. 100:267–280.

Macarthur, D. G., T. A. Manolio, D. P. Dimmock, H. L. Rehm, J. Shendure, G. R. Abecasis, et al. 2014. Guidelines for investigating causality of sequence variants in human disease. Nature 508:469–476.

Majounie, E., A. E. Renton, K. Mok, E. G. P. Dopper, A. Waite, S. Rollinson, et al. 2012. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. Lancet Neurol. 11:323–330.

Marangi, G., and B. J. Traynor. 2015. Genetic causes of amyotrophic lateral sclerosis: new genetic analysis methodologies entailing new opportunities and challenges. Brain Res. 1607:75–93.

McCann, E. P., K. L. Williams, J. A. Fifita, I. S. Tarr, J. O'connor, D. B. Rowe, et al. 2017. The genotype-phenotype landscape of familial amyotrophic lateral sclerosis in Australia. Clin. Genet. https://doi.org/10.1111/cge.12973.

McCombe, P. A., N. R. Wray, and R. D. Henderson. 2016. Extra-motor abnormalities in amyotrophic lateral sclerosis: another layer of heterogeneity. Expert Rev. Neurother. 1–17.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20: 1297–1303.

Morgan, S., M. Shoai, P. Fratta, K. Sidle, R. Orrell, M. G. Sweeney, et al. 2015. Investigation of next-generation sequencing technologies as a diagnostic tool for amyotrophic lateral sclerosis. Neurobiol. Aging 36:1600.e1605-1608.

Oakes, J. A., M. C. Davies, and M. O. Collins. 2017. TBK1: a new player in ALS linking autophagy and neuroinflammation. Mol. Brain 10:5.

Orrù, S., E. Manolakos, N. Orrù, H. Kokotas, V. Mascia, C. Carcassi, et al. 2012. High frequency of the TARDBP p.Ala382Thr mutation in Sardinian patients with amyotrophic lateral sclerosis. Clin. Genet. 81:172–178.

Ozoguz, A., O. Uyan, G. Birdal, C. Iskender, E. Kartal, S. Lahut, et al. 2015. The distinct genetic pattern of ALS in Turkey and novel mutations. Neurobiol. Aging 36:1764.e1769-1718.

Quadri, M., G. Cossu, V. Saddi, E. J. Simons, D. Murgia, M. Melis, et al. 2011. Broadening the phenotype of TARDBP mutations: the TARDBP Ala382Thr mutation and Parkinson's disease in Sardinia. Neurogenetics 12:203–209.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, et al. 2011a. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72:257–268.

Renton, A. E., A. Chio, and B. J. Traynor. 2014. State of play in amyotrophic lateral sclerosis genetics. Nat. Neurosci. 17:17–23.

Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. 17:405–424.

Roggenbuck, J., A. Quick, and S. J. Kolb. 2016. Genetic testing and genetic counseling for amyotrophic lateral sclerosis: an update for clinicians. Genet. Med. 19:267–274 doi:10.1038/gim.2016.107.

Rosen, D. R., T. Siddique, D. Patterson, D. A. Figlewicz, P. Sapp, A. Hentati, et al. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 362:59–62.

Simons, C., L. B. Griffin, G. Helman, G. Golas, A. Pizzino, M. Bloom, et al. 2015. Loss-of-function alanyl-tRNA synthetase mutations cause an autosomal-recessive early-onset epileptic encephalopathy with persistent myelination defect. Am. J. Hum. Genet. 96:675–681.

Stenson, P. D., M. Mort, E. V. Ball, K. Shaw, A. Phillips, and D. N. Cooper. 2014. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. Hum. Genet. 133:1–9.

Sun, Y., R. Almomani, G. J. Breedveld, G. W. Santen, E. Aten, D. J. Lefeber, et al. 2013. Autosomal recessive spinocerebellar ataxia 7 (SCAR7) is caused by variants in TPP1, the gene involved in classic late-infantile neuronal ceroid lipofuscinosis 2 disease (CLN2 disease). Hum. Mutat. 34:706–713.

Tamaoka, A., M. Arai, M. Itokawa, T. Arai, M. Hasegawa, K. Tsuchiya, et al. 2010. TDP-43 M337V mutation in familial amyotrophic lateral sclerosis in Japan. Int. Med. (Tokyo, Japan) 49:331–334.

Vajda, A., R. L. Mclaughlin, M. Heverin, O. Thorpe, S. Abrahams, A. Al-Chalabi, et al. 2017. Genetic testing in ALS: a survey of current practices. Neurology 88:991–999.

Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 38: e164.

Wu, C. H., C. Fallini, N. Ticozzi, P. J. Keagle, P. C. Sapp, K. Piotrowska, et al. 2012. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature 488:499–503.

Xi, Z., L. Zinman, D. Moreno, J. Schymick, Y. Liang, C. Sato, et al. 2013. Hypermethylation of the CpG Island near the G (4)C(2) Repeat in ALS with a C9orf72 expansion. Am. J. Hum. Genet. 92:981–989.

Supporting Information

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

Table S1. Gene list.
Table S2. Known ALS variants identified in patients that were likely causal.
Table S3. Previously reported ALS-associated variants found in Brisbane ALS patients at a minor allele frequency which renders them unlikely to be causal.
Table S4. Novel rare damaging variants in ALS genes identified in ALS cases of unknown consequence likely comprising both true and false positives.
Table S5. Novel rare damaging variants in similar disease genes identified in ALS cases.
Table S6. Recent approaches used to genetically screen ALS patients.
Figure S1. Boxplots of the methylation signal in cases and controls for the *C9orf72* probes and DNA methylation levels in *C9orf72* gene region.
Figure S2. (A) Boxplots of the methylation signal in cases and controls for the *SOD1* probes and DNA methylation levels in the *SOD1* gene region.
Appendix S1. Materials, methods and results.