Genome-wide Analyses Identify KIF5A as a Novel ALS Gene

DOI: 10.1016/j.neuron.2018.02.027

Document Version
Accepted author manuscript

Link to publication record in Manchester Research Explorer

Citation for published version (APA):

Published in:
Neuron

Citing this paper
Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights
Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy
If you believe that this document breaches copyright please refer to the University of Manchester’s Takedown Procedures [http://man.ac.uk/04Y6Bo] or contactuml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.

Download date: 12. Aug. 2022
Genome-wide Analyses Identify KIF5A as a Novel ALS Gene

To identify novel genes associated with ALS, we undertook two lines of investigation. We carried out a genome-wide association study comparing 20,806 ALS cases and 59,804 controls. Independently, we performed a rare variant burden analysis comparing 1,138 index familial ALS cases and 19,494 controls. Through both approaches, we identified kinesin family member 5A (KIF5A) as a novel gene associated with ALS. Interestingly, mutations predominantly in the N-terminal motor domain of KIF5A are causative for two neurodegenerative diseases, hereditary spastic paraplegia (SPG10) and Charcot-Marie-Tooth Type 2 (CMT2). In contrast, ALS associated mutations are primarily located at the C-terminal cargo-binding tail domain and patients harboring loss of function mutations displayed an extended survival relative to typical ALS cases. Taken together, these results broaden the phenotype spectrum resulting from mutations in KIF5A and strengthen the role of cytoskeletal defects in the pathogenesis of ALS.
INTRODUCTION

Amyotrophic lateral sclerosis (ALS, OMIM #105400) is a neurodegenerative disorder clinically characterized by rapidly progressive muscle weakness and death due to respiratory failure, typically within two to four years of symptom onset (van Es et al., 2017). Although ALS is perceived as being rare, approximately 6,000 Americans die annually from the condition (Hirtz et al., 2007). Furthermore, the number of ALS cases across the globe will increase to nearly 400,000 in 2040, predominantly due to aging of the population (Arthur et al., 2016). This increase is anticipated to place an enormous socioeconomic burden on global healthcare systems, in particular because the annual healthcare cost per patient with ALS is among the highest for any neurological disease (Gladman and Zinman, 2015).

Approximately 10% of ALS display a family history (FALS) whereas the remaining 90% of ALS cases are sporadic (SALS) in nature. Driven in large part by advances in genotyping and sequencing technology, the genetic etiology of two-thirds of familial cases and about 10% of sporadic ALS cases is now known (Chia et al., 2018; Renton et al., 2014). Mutations in SOD1 were the first identified cause of ALS (Rosen et al., 1993) contributing to ~20% of FALS and ~2% of SALS. More recently, pathogenic hexanucleotide repeat expansions located within the first intron of the C9orf72 gene on chromosome 9p21 were identified as the most common cause of both FALS (~40%) and SALS (~7%) (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Interestingly, this repeat expansion contributes to ~10% of all FTD cases thus genetically explaining much of the overlap between these clinical syndromes (Majounie et al., 2012). As a result of these major discoveries, there are several ongoing efforts towards directed silencing of these mutant genes which could result in a therapeutic treatment for up to 10% of all ALS cases and for a similar portion of FTD cases.
In addition to the insights provided by each novel ALS gene, the collective knowledge gained from genetic factors provides a more comprehensive understanding of the interacting pathways underlying motor neuron degeneration. For example, the identification of ALS genes has revealed at least three pathways believed to contribute to the development of ALS: (1) RNA metabolism (based on the observation of mutations in \textit{C9orf72}, \textit{TDP-43}, \textit{FUS}, \textit{HNRNPA1}, and \textit{MATR3}); (2) protein homeostasis (\textit{UBQLN2}, \textit{VCP}, \textit{OPTN}, \textit{VAPB}); (3) cytoskeletal dynamics (\textit{PFN1}, \textit{TUBA4A}, \textit{DCTN1}) (Chia et al., 2018; Robberecht and Eykens, 2015; Taylor et al., 2016). Understanding the mechanisms leading to disease pathogenesis again provides targets for therapeutic intervention that may be applicable to all forms of ALS.

Due to the decreased accessibility of multiple affected family members with unknown genetic etiology, there has been an increased focus on the identification of ALS associated genes with moderate to low impact. Despite their lower effect, such genes continue to provide valuable insight into ALS pathogenesis. For example, the product of the risk factor \textit{TBK1} is known to interact with the product of ALS associated gene \textit{OPTN}, further solidifying the role of autophagy and protein homeostasis in disease development (Cirulli et al., 2015; Freischmidt et al., 2015; Maruyama et al., 2010; Morton et al., 2008). Similarly, the risk factor \textit{NEK1}, identified through a rare variant burden analysis of index FALS (i.e., one affected sample per family), is a known binding partner of \textit{C21orf2}, an ALS risk factor found through genome-wide association studies (GWAS) (Cirulli et al., 2015; Kenna et al., 2016; Malovannaya et al., 2011; van Rheenen et al., 2016). The interaction of these two proteins is required for efficient DNA damage repair (Fang et al., 2015), a pathway which is becoming increasingly implicated as a contributing factor in ALS and other neurodegenerative diseases (Coppedè and Migliore, 2015; Lopez-Gonzalez et al., 2016; Madabhushi et al., 2014; Wang et al., 2013).
RESULTS

Genome-wide Association Studies Identify KIF5A as a Novel ALS Associated Gene

To identify new susceptibility loci operating in ALS, we undertook a large-scale GWAS involving 12,663 patients diagnosed with ALS and 53,439 control subjects (Table S1, S2). Our data were then incorporated into a meta-analysis with a recently published GWAS involving 12,577 ALS cases and 23,475 control subjects (van Rheenen et al., 2016). After imputation and quality-control measures (see Methods, Figure S1 for the workflow and Figure S2 for the multidimensional scaling plot), 10,031,630 genotyped and imputed variants from 20,806 ALS cases and 59,804 control samples were available for association analysis (Figure 1A). Quantile-quantile plots did not show evidence of significant population stratification ($\lambda_{1000} = 1.001$, Figure S3). Single nucleotide polymorphisms (SNPs) achieving genome-wide significance ($P < 5.0 \times 10^{-8}$) are listed in Table 1, Table S3 and suggestive loci with SNPs associated at $P < 5.0 \times 10^{-7}$ are listed in Table S4.

Our analysis revealed five previously identified loci that achieved genome-wide significance (loci including TNIP1, C9orf72, TBK1, UNC13A, C21orf2) (Benyamin et al., 2017; Laaksovirta et al., 2010; Shatunov et al., 2010; van Es et al., 2009; van Rheenen et al., 2016). In addition, we observed a strong association signal for five SNPs in linkage disequilibrium on chromosome 12q14.1 that reached genome-wide statistical significance (Table 1, Figure 1B) spanning a region of several hundred kilobases. Of the five SNPs, two of them resided in close proximity to each other within a large intergenic region and two in proximity to short-chain dehydrogenase/reductase family 9C member 7 (SDR9C7), a gene expressed primarily in skin. However, one SNP (rs113247976) results in a p.Pro986Leu coding change within the kinesin family member 5A (KIF5A) gene ($P = 6.4 \times 10^{-10}$, OR = 1.38, 95% CI = 1.24-1.53). The case:control allele frequencies for the combined discovery cohort were
2.07%:1.55% and genotype counts were 5 : 529 : 12,043 to 7 : 786 : 22,682 (homozygotes alternative allele : heterozygotes : homozygous reference allele, Figure 2). Calculations based on our cohort size as well as the OR and allele frequency of rs113247976 result in a ~99.5% power to detect this as an ALS associated SNP.

**Rare Variant Burden Analysis Identifies KIF5A as an ALS gene**

In an independent line of investigation, we attempted to identify novel ALS genes through exome-wide rare variant burden analysis (RVB). In brief, RVB compares the frequency of variants within each gene below a user defined frequency threshold in a case-control cohort. As the last two ALS associated genes identified by this methodology (TBK1, NEK1) displayed an increased frequency of loss of function (LOF) variants, we focused our initial analysis on such variants (consisting of nonsense and predicted splice-altering) (Cirulli et al., 2015; Freischmidt et al., 2015; Kenna et al., 2016).

Towards this end, we performed RVB testing for association of LOF variants in a cohort of 1,138 index FALS cases and 19,494 controls, after applying quality control filters (Experimental Methods, Figure S4, Table S5, S6). Genes displaying $P < 5.0 \times 10^{-4}$ are shown in Table 2. The previously identified ALS genes, TBK1 ($P = 5.58 \times 10^{-7}$, OR = 15.11, 95% CI = 5.81-38.69) and NEK1 ($P = 1.68 \times 10^{-6}$, OR = 6.64, 95% CI = 3.32-12.51), yielded strong associations with ALS reaching exome-wide significance (Figure 3). In addition, we observed a single novel gene reaching exome-wide significance, KIF5A ($P = 5.55 \times 10^{-7}$; OR = 32.07, 95% CI = 9.05-135.27). Within this gene, we observed 6 LOF variants in our 1,138 cases (0.53%) compared to 3 such variants in our comparison cohort of 19,494 controls (0.015%, Table 2). There was no evidence of genomic inflation ($\lambda$ =0.93, Figure S5), sequencing center or other sub-cohort bias (Figure S6), or call rate bias (Figure S7) in our analysis. Of the index FALS cases
carrying *KIF5A* LOF mutations, we obtained DNA from two siblings of the proband carrying a c.2993-3C>T, exon 27 - 5' splice junction variant, and from a sibling of a different proband carrying a c.3020+2T>A, exon 27 - 3' splice junction variant. These variants segregated with disease within each of these families. Sanger sequencing was performed for 8 of the 9 LOF variant containing samples and was validated in every case.

Interestingly, when we investigated the location of the six ALS associated variants present in *KIF5A*, all occurred within a 34 bp stretch of DNA and were predicted to effect splicing of exon 27 that encodes amino acids 998-1007 (*Table 3, Figure 4A*). Five of the six variants were located on sequential base pairs on the 3’ end of the exon, whereas one was located 5’ end of the exon. We used the application ASSEDA (Automated Splice Site and Exon Definition Analyses) to predict any mutant mRNA splice isoforms resulting from these variants (Tompson et al., 2007). This algorithm was chosen as it is known to have high performance in splice prediction (Caminsky et al., 2014). ASSEDA predicted a complete skipping of exon 27 for all variants, yielding a transcript with a frameshift at coding amino acid 998, the deletion of the normal C-terminal 34 amino acids of the cargo-binding domain, and the extension of an aberrant 39 amino acids to the C-terminus (*Table 3, Figure 4B, 4C*). The presence of transcripts with skipped exon 27 was demonstrated by performing RT-PCR in two patients carrying exon 27 - 3' splice junction variants (c.3020+2T>A and c.3020+1G>A) using RNA from lymphoblasts and peripheral blood mononuclear cells, respectively. This splice form was not detected in four control lines (*Figure 4D*). Sequence analysis of the smaller RT-PCR products obtained from the patient cells confirmed the exon 26-28 splicing event. Material for RT-PCR was not available for any other patient carrying a *KIF5A* LOF variant.
Our initial RVB was restricted to single nucleotide variants due to the limited sensitivity and comparatively high false positive rates associated with identifying small insertions and deletions (indels) within exome sequencing data (Fang et al., 2014). Based on our discovery of increased LOF variants within KIF5A, we re-evaluated this region for the presence of indels. Our analysis revealed two (0.026%) indels within our cohort of 1,138 FALS cases, compared to zero (0%) indels among 19,494 control samples. Both of these indels (p.Asp996fs, p.Asn999fs) resulted in a frameshift of the KIF5A protein coding sequence, and were located close to the splice junction variants that we previously observed to cause skipping of exon 27 resulting in a frameshift at amino acid 998 (Table 3). Sanger sequencing confirmed the presence of both indels. Combining the results of the single nucleotide and indel variant analysis yielded a highly significant $P$ of $3.8 \times 10^{-9}$ ($OR = 41.16, 95\% CI = 12.61-167.57$).

We failed to detect any signals of RVB association for rare missense variants across KIF5A or within any sub-domain of the gene (Table S8).

**Replication Analysis of rs113247976 and LOF Variants in KIF5A**

Given the strong signal of the missense variant identified by our GWAS (p.Pro986Leu, rs113247976) and its close proximity to the LOF variants identified by our RVB analysis (amino acids 996-999), we attempted to replicate its association with ALS by analyzing additional cohorts. To accomplish this, we evaluated this variant in a cohort of 4,159 ALS cases and 18,650 controls that were non-overlapping with our GWAS discovery analysis (Methods, Figure S8). This included non-overlapping samples from our RVB analysis (673 FALS, 17,696 controls). Analysis of the cohort revealed an allele frequency of 1.78% in cases and 1.32% in controls (rs113247976, $P = 3.82 \times 10^{-4}$, $OR = 1.42, 95\% CI = 1.17-1.70$), thereby replicating the association of the original GWAS. A meta-analysis of the GWAS and replication cohort ($n = 24,965$ cases, $78,454$ controls) yielded a highly significant $P$ of $7.09 \times 10^{-13}$ ($OR = 1.39, 95\%$
CI = 1.33-1.45, Figure 2). These results support the association of $KIF5A$ p.Pro986Leu with ALS. However, at this point we cannot definitely state that the missense variant is the primary risk factor, as we cannot rule out other variants in linkage disequilibrium.

We next performed mutational screening of $KIF5A$ in an additional cohort of 9,046 ALS cases that had not been included in our original RVB analysis. This revealed three additional carriers of C-terminal variants. One sporadic patient harbored an exon 26 frameshift mutation (p.Asn997fs) and a second sporadic patient harbored an exon 27 splice altering mutation (c.2993-1G>A, Table 3). The third patient carried a p.Arg1007Lys (c.3020G>A) mutation and had a familial history of ALS. This mutation was also observed in a FALS patients from our RVB analysis, however, a comparison of 240,715 common variant sites between the two patients failed to reveal a familial relationship (Experimental Methods). Additionally, one patient was observed to carry a predicted splice altering variant proximal to exon 3 (c.291+5G>A). However, this variant was not supported as creating an aberrant transcript by ASSEDA. The cohort used for this analysis was comprised mainly frequency of sporadic ALS cases. LOF variants were not observed in a follow up panel of 1,955 controls. Comparison of the LOF variants in sporadic patients (2/9,046 cases, 0.022%) with either the 1,955 replication controls or all controls analyzed in this study (21,449 controls) both yielded insignificant $P$ values (0.868 and 0.423, respectively). Interestingly, the frequency of LOF variants in sporadic cases is lower than that observed in our original FALS cohort (0.703%), suggesting that $KIF5A$ LOF variants display a high penetrance. Furthermore, the rate of LOF variants reported in the Exome Aggregation Consortium (ExAC) database is lower than we observed in the control samples used in our discovery cohort (0.007% versus 0.015%).

**ALS-Associated Mutations in $KIF5A$ are Distinct from SPG10/CMT2 Mutations**
Missense mutations within \textit{KIF5A} are a known cause of hereditary spastic paraparesis (spastic paraplegia type 10, autosomal dominant; OMIM #604187) and of Charcot-Marie-Tooth disease Type 2 (CMT2) (Crimella et al., 2011; Jennings et al., 2017; Liu et al., 2014; Reid et al., 2002). Although SPG10 and CMT2 share clinical features with ALS, a careful examination of the clinical records of the ALS cases with LOF mutations in \textit{KIF5A} ruled out misdiagnosis. Furthermore, we detected no variants previously associated with SPG10 or CMT2 in our FALS cohort (Liu et al., 2014).

To further elucidate genotype-phenotype relationships, we evaluated the location of mutations within \textit{KIF5A}. Interestingly, mutations contributing to SPG10 and to CMT2 are almost exclusively to be missense changes and are located in the N-terminal motor domain (amino acids 9-327) of KIF5A (Figure 5). In contrast, the mutations identified as contributing to ALS are found predominantly in the C-terminal cargo binding region of KIF5A (amino acids 907-1032) with the highly penetrant FALS mutations showing LOF. These results indicate that the functional domain mutated in KIF5A dictates the clinical phenotype, resulting in distinct yet overlapping neurodegenerative diseases.

\textbf{Patients with \textit{KIF5A} LOF Mutations Display Younger Age at Onset and Longer Survival}

To establish the existence of any commonalities between patients with LOF mutations in the C-terminal region of KIF5A, we evaluated their clinical phenotype. Cases with LOF mutations exhibited a median age of onset at 46.5 years (n = 19, Table S7). This is lower than the age of onset reported for ALS in epidemiological studies (65.2 years, interquartile range 56.0 – 72.2) (ALSGEN Consortium et al., 2013). Interestingly, we also observed an increased disease duration (survival) in patients harboring these LOF mutations. The median survival time of ALS patients is 20 – 36 months (ALSGEN Consortium et al., 2013). In contrast, cases with LOF mutations exhibited a median survival of nearly 10 years (117
months, n = 17, Table S7). ALS patients with symptom onset before 40 years of age have been shown to have longer survival, often exceeding 10 years (Chio et al., 2009). In contrast, patients with uncomplicated types of hereditary spastic paraparesis and CMT2 display a normal life expectancy (Patzkó and Shy, 2011).

**DISCUSSION**

We previously identified *KIF5A* as a candidate gene for ALS in our prior study that lacked the power to draw a definitive conclusion (Kenna et al., 2016). *KIF5A* was also a candidate ALS gene in a previous GWAS, though it similarly failed to reach genome-wide significance (McLaughlin et al., 2017; van Rheenen et al., 2016) as well as a single gene study selected based on the *a priori* knowledge of its role in HSP/CMT2 and cytoskeletal function (Brenner et al., 2018). Here, we have confirmed *KIF5A* as an ALS-associated gene with genome-wide significance through two independent approaches. By performing a GWAS involving ~80,000 samples, in addition to replicating five previously published loci, as well as the previously reported locus *SCFD1* using a linear mixed model analysis (data not shown), we identified a missense variant within the *KIF5A* gene that reached genome-wide significance for association with ALS risk. It should be stated though that, as with all GWAS, we cannot rule out that other variants in linkage disequilibrium represent the primary risk factor. In an independent line of investigation, we applied RVB analysis to exome sequencing of ~21,000 samples and identified an exome-wide significant association between FALS risk and rare *KIF5A* LOF variants. Analyses of *KIF5A* in independent replication cohorts confirmed our initial finding for both the p.Pro986Leu variant and revealed three additional carriers of LOF variants in 9,046 ALS cases. Taken together our results indicate that the p.Pro986Leu *KIF5A* variant may represent a relatively common, but low penetrance
risk allele for ALS, while LOF variants constitute rare, but high penetrance risk factors.

Kinesins are microtubule-based motor proteins involved in intracellular transport of organelles within eukaryotic cells. In mammals, there are three heavy chain isoforms of KIF5: KIF5A, KIF5B and KIF5C (Miki et al., 2001). The three proteins homo- and heterodimerize through their coiled-coiled stalk domain, and create a complex with two kinesin light chains via binding to the tail domain (Hirokawa et al., 1989). All three KIF5 genes are expressed in neurons (Kanai et al., 2000) and function to transport many cargos by binding to distinct adaptor proteins.

The central role of kinesins in axonal transport leads us to speculate that mutations in KIF5A cause disease by disrupting axonal transport. Indeed, defects in axonal transport are a common observation in ALS patients and are already known to directly contribute to motor neuron degeneration pathogenesis (Chevalier-Larsen and Holzbaur, 2006; Hirokawa et al., 2010; Millecamps and Julien, 2013). KIF5 mediates the transport of granules containing both RNA and RNA binding proteins within neuronal dendrites and axons (Kanai et al., 2004). Among these cargos are the ALS-associated proteins FUS and hnRNPA1 (Guo et al., 2017; Kim et al., 2013; Kwiatkowski et al., 2009; Vance et al., 2009). Similarly, KIF5 mediates the transport of VAPB through the adaptor protein protrudin (Matsuzaki et al., 2011), and mutations in the VAPB gene have been identified in ALS and late-onset spinal muscular atrophy (Nishimura et al., 2005; 2004). KIF5 is responsible for the axonal transport of neurofilaments (Wang and Brown, 2010) and KIF5A knockout mice display abnormal transport of neurofilaments (Xia et al., 2003). Abnormal accumulation of neurofilaments are a pathological hallmark of ALS and rare mutations in neurofilament heavy polypeptide (NEFH) are associated with ALS (Al-Chalabi et al., 1999).
KIF5 also contributes to the transport of mitochondria (Kanai et al., 2000; Tanaka et al., 1998) and motor neurons derived from KIF5A−/− mice display transport deficits and reduced survival (Karle et al., 2012). Impaired transport and dysfunction of mitochondria represent another common hallmark observed in ALS patients (Chevalier-Larsen and Holzbaur, 2006; Guo et al., 2017; Palomo and Manfredi, 2015; Smith et al., 2017). KIF5 also contributes to the transport of AMPA-type (Heisler et al., 2014; Setou et al., 2002) and GABA$_A$ receptors (Nakajima et al., 2012). In keeping with reported ALS genes such as NEK1 (Thiel et al., 2011) and PFN1 (Wu et al., 2012), modulation of KIF5A expression has been shown to influence the formation of neurite like membrane protrusions (Matsuzaki et al., 2011). Given its critical interactions with the cytoskeleton, the identification of KIF5A mutations further extends the list of cytoskeletal related proteins implicated in ALS pathogenesis, such as PFN1, TUBA4A, NEFH and peripherin (Al-Chalabi et al., 1999; Gros-Louis, 2004; Smith et al., 2014; Wu et al., 2012).

An important question raised by the current study is why variation within the C-terminal cargo binding domain is associated with ALS, while missense variations of the N-terminal motor domain are associated with hereditary spastic paraparesis and Charcot-Marie-Tooth, type 2. Missense mutations within this latter domain have been shown to affect microtubule binding and/or ATP hydrolysis, resulting in a defective KIF5A-mediated anterograde transport of cargo along dendrites and axons. This, in turn, leads to the axonal retrograde degeneration observed both in hereditary spastic paraparesis and Charcot-Marie-Tooth, type 2, two length-dependent axonopathies (Ebbing et al., 2008). In contrast, the primary cellular lesion in ALS is believed to occur within motor neuron cell bodies, where cytoplasmic protein aggregates are consistently observed, and to propagate anterograde along neurites. We anticipate
that LOF variants within the C-terminal domain of KIF5A will disrupt binding with specific cargo proteins. This is supported by a study in zebrafish in which truncation of the C-terminal resulted in a dramatic disruption of axonal localization of mitochondria (Campbell et al., 2014). One possible mechanism is that disruption of binding to cargo may possibly lead to their accumulation and seed aggregation within the cell body resulting in a deficiency at neurite terminals. Deficiency in KIF5A expression and cargo binding has been associated with accumulation of phosphorylated neurofilaments and amyloid precursor protein within neuronal cell bodies, and subsequent neurodegeneration, in patients with multiple sclerosis (Hares et al., 2016). While differences in KIF5A kinetics and KIF5A interactions constitute one possibility to explain the phenotypic heterogeneity, it is also possible C-terminal and N-terminal variants act through a common mechanism, but that a difference in the relative extent of loss or gain of function toxicities leads to milder (i.e. hereditary spastic paraplegia, Charcot-Marie-Tooth, type 2) or more severe (i.e. ALS) phenotypes.

ACKNOWLEDGEMENTS

Author Contributions:


Conflict of Interest Disclosures: None of the authors report any conflicts of interest. J.D.Be. is a consultant to Neuraltus Pharmaceuticals and Denali Therapeutics, and held a research fellow position.
funded by Voyager Therapeutics. M.C. has been a consultant for Eli Lilly and Company, Mitsubishi Tanabe Pharma America (MT Pharma America), Denali Therapeutics, Karyopharm Therapeutics and Cytokinetics. S.A.G. has served as a consultant and received research support from Cytokinetics. O.H. has received speaking honoraria from Novartis, Biogen Idec, Sanofi Aventis and Merck-Serono and has been a member of advisory panels for Biogen Idec, Allergen, Ono Pharmaceuticals, Novartis, Cytokinetics and Sanofi Aventis. O.H. serves as Editor-in-Chief of Amyotrophic Lateral Sclerosis and Frontotemporal Dementia. L.H.v.d.B. serves on scientific advisory boards for the Prinses Beatrix Spierfonds, Thierry Latran Foundation, Biogen and Cytokinetics. Serves on the editorial board of Amyotrophic Lateral Sclerosis And Frontotemporal Degeneration and The Journal of Neurology, Neurosurgery, and Psychiatry. J.H.V. reports that his institute received consultancy fees from Vertex Pharmaceuticals. A.C. serves on scientific advisory boards for Biogen Idec, Cytokinetics, Italfarmaco, and Neuraltus.

**Funding/Support:** P.V.D. is a senior investigator of FWO-Vlaanderen. Project MinE Belgium has been supported by ALS liga België, Flanders Innovation & Enterpreneurship (IWT grant Project MinE), the Belgian National Lottery and a grant from Opening the Future Fund (KU Leuven). W.v.R. is supported through the E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders and ERC (grant agreement n° 340429). Additional funding support includes NINDS R35 NS097261 (R.R.) and P01NS084974 (R.R. and K.B.B.). A.N.B. thanks Suna and Inan Kirac Foundation for its generous support throughout this study. Funding for this work was provided by the Heaton-Ellis Trust, The Middlemass Family, Motor Neurone Disease Association, Medical Research Council, Medical Research Foundation, The Psychiatry Research Trust of the Institute of Psychiatry, Guy’s and St. Thomas’ Charity, the Wellcome Trust, and the Noreen Murray Foundation (C.E.S). This work was also supported
by the UK Dementia Research Institute which is funded by the Medical Research Council, Alzheimer’s Society and Alzheimer’s Research UK (C.E.S.). The salary for B.N.S was funded by the Medical Research Foundation (MRF) (MRF-060-0003-RG-SMITH). P.C.S. was supported through the auspices of Dr. H. Robert Horvitz (Massachusetts Institute of Technology), an Investigator of the Howard Hughes Institute. Support from the Department of Veterans Affairs and NIH (P30AG13846) to N.W.K. I.P.B. is supported by the Motor Neurone Disease Research Institute of Australia and the National Health and Medical Research Council of Australia (1107644, 1095215). Research support from NIH/NIEHS (K23ES027221), the ALS Association, Target ALS, and Cytokinetics was provided to S.A.G. M.C. was awrded funding from ALS Finding a Cure. N.T., C.Ti., C.G., V.S. and J.E.L. received research support from AriSLA - Fondazione Italiana di Ricerca per la SLA (grants EXOMEFALS and NOVALS) and the Italian Ministry of Health (grant GR-2011-02347820 - IRisALS). R.L.M. was supported by Science Foundation Ireland and the MND Association of England, Wales and Northern Ireland. O.H. is funded by the Health Research Board Clinician Scientist Programme and Science Foundation Ireland. P.J.S. is supported as an NIHR Senior Investigator (NF-SI-0512-10082). P.J.S. and J.K. are supported by the Sheffield NIHR Biomedical Research Centre for Translational Neuroscience (IS-BRC-1215-20017). A.C. receives research support from the Italian Ministry of Health (Ricerca Finalizzata), Regione Piemonte (Ricerca Finalizzata), University of Turin, Fondazione Vialli e Mauro onlus, and the European Commission (Health Seventh Framework Programme). R.B. received funding support from NINDS/NS061867 and Target ALS. R.H.B.J. received funding from the Angel Fund, Project ALS/P2ALS, and the ALS Therapy Alliance. Additional funding was provided by the NIH/NINDS (R01NS073873, J.E.L.), the American ALS Association (N.T., V.S. C.E.S., R.H.B.J., J.E.L.), and the MND Association (N.T., V.S., C.E.S., J.E.L.).
The sequencing activities at NYGC were supported by the ALS Association (ALSA) and TOW Foundation. The CReATe consortium (U54NS092091) is part of Rare Diseases Clinical Research Network (RDCRN), an initiative of the Office of Rare Diseases Research (ORDR), NCATS. This consortium is funded through collaboration between NCATS, and the NINDS. Additional support is provided by the ALS Association (17-LGCA-331). The InCHIANTI study baseline (1998–2000) was supported as a “targeted project” (ICS110.1/RF97.71) by the Italian Ministry of Health and in part by the U.S. National Institute on Aging (Contracts: 263 MD 9164 and 263 MD 821336); the InCHIANTI Follow-up 1 (2001–2003) was funded by the U.S. National Institute on Aging (Contracts: N.1-AG-1-1 and N.1-AG-1-2111); the InCHIANTI Follow-ups 2 and 3 studies (2004–2010) were financed by the U.S. National Institute on Aging (Contract: N01-AG-5-0002). This work was supported in part by the Intramural Research Programs of the US National Institutes of Health (NIH), National Institute on Aging (Z01-AG00949-02); by the National Institute of Neurological Disorders and Stroke; and by Merck & Co., Inc. The work was also supported by the Center for Disease Control and Prevention, the Muscular Dystrophy Association, Microsoft Research, the Packard Center for ALS Research at Johns Hopkins, the ALS Association, UK MND Association, Medical Research Council (MRC) UK, Wellcome Trust/MRC Joint Call in Neurodegeneration Award, MRC Neuromuscular Centre, UK National Institute for Health Research Biomedical Research Unit, Italian Health Ministry (Ricerca Sanitaria Finalizzata 2007), Fondazione Vialli e Mauro Onlus, Compagnia di San Paolo, European Community’s Health Seventh Framework Programme (FP7/2007-2013) under grant agreements 259867.

**Role of the Sponsors:** The sponsors did not participate in the design and conduct of the study; collection, management, analysis or interpretation of the data; or preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.
**Additional contributions:** We thank the patients and research subjects who contributed samples for this study. This study used DNA samples and clinical data from the Target ALS Human Postmortem Tissue Core, the NINDS Repository at Coriell, the North East ALS (NEALS) Consortium Biorepository, the New York Brain Bank-The Taub Institute, Columbia University, Department of Veterans Affairs Biorepository Brain Bank (grant #BX002466; C.B.B.), the Baltimore Longitudinal Study on Aging at the Johns Hopkins School of Medicine, the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland and the Australian MND DNA Bank. We also thank Crystal Pacut, Blake Swihart, and Jayna Duell, RN for assistance in study coordination (S.A.G.). This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Maryland (http://biowulf.nih.gov) and the Massachusetts Green High Performance Computing Center at the University of Massachusetts Medical School (https://www.mghpcc.org). This study also used genotype and clinical data from the Wellcome Trust Case Control Consortium.
REFERENCES


Salvi, E., Kutilik, Z., Glorioso, N., Benaglio, P., Frau, F., Kuznetsova, T., Arima, H., Hoggart, C.,


Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B.,


MAIN FIGURE TITLES AND LEGENDS

Figure 1. Identification of association between KIF5A locus and ALS risk through GWAS. (A) Manhattan plot showing $P$ values from the discovery set GWAS. Analysis of a combined set of 20,806 cases and 59,804 controls is shown. The dashed red line denotes the threshold for genome-wide significance after multiple test correction ($P < 5.0 \times 10^{-8}$). Five previously reported ALS associated loci are labeled in grey and one novel loci, containing the KIF5A gene, is labeled in black. (B) Regional association plot of the KIF5A locus. Recombination rates are from HapMap phase 2 European ancestry samples. The $R^2$ pattern is based on the rs113247976 SNP using 85 European ancestry samples (CEU) from the November 2010 release of the 1000 Genomes Project dataset. $R^2$ of the p.Pro986Leu (rs113247976) with additional SNPs achieving genome-wide significance was 0.544 (rs117027576), 0.544 (rs118082508), 0.741 (rs116900480), and 0.347 (rs142321490).

Figure 2. Discovery and replication for the association of the KIF5A p.Pro986Leu (rs113247976) variant with ALS. Analysis of the p.Pro986Leu (rs113247976) variant within each of the described cohorts is shown. Allelic association for all subcohorts were analyzed by logistic regression followed by a fixed-effects meta-analysis. The Forest plot (right) displays the distribution of OR estimates across study cohorts with the vertical dotted line denoting the OR estimated under the meta-analysis.

Figure 3. Identification of association between KIF5A and ALS risk through rare variant burden analysis of exome sequencing. Manhattan plot showing gene-level $P$ values from an exome-wide rare variant burden analysis. Analyses of 1,138 index FALS cases versus 19,494 controls were restricted to rare LOF variants (splice altering/nonsense, MAF $< 0.001$). A minimum of 3 LOF gene variants were
required for analysis. The dashed red line denotes the threshold for exome-wide significance after correction for 11,472 genes (4.36x10^6). Previously reported (grey) and novel (black) genes exhibiting a significant excess of rare LOF variants in patients are shown.

Figure 4. ALS associated loss of function variants of KIF5A disrupt C-terminal sequence by inducing skipping of exon 27. (A) Single nucleotide variants (SNVs) within KIF5A identified in ALS patients are clustered at the 5’ and 3’ splice junctions of exon 27. The consensus splice sequence is shown. (B) ALS associated SNVs are predicted to induce skipping of exon 27 and result in an aberrant mRNA transcript. (C) The skipping of exon 27 of KIF5A yields an out-of-frame and extended disrupted C-terminal peptide sequence. The amino acids in red signify the divergence from the normal protein. (D) RT-PCR was performed using RNA derived from ALS patients with the indicated LOF variant or without (controls) using primers to either amplify both wild-type (155 bp) and mutant (127 bp) splice forms or specifically the mutant splice form (80 bp, right panel). The arrow represents the position of the mutant specific product. The tick marks represent 200 bp (upper) and 100 bp (lower) markers.

Figure 5. KIF5A ALS mutations show distinct localization from missense mutations previously associated with SPG10 and CMT2. Causative mutations for SPG10 and CMT2 described within the literature (Crimella et al., 2011; Jennings et al., 2017; Liu et al., 2014; Reid et al., 2002) and ALS associated mutations identified within this study are shown. As illustrated, mutations causative for SPG10/CMT2 are predominantly missense changes located in the N-terminal motor domain. In contrast, ALS mutations are primarily located at the C-terminal motor domain and are LOF.
Table 1. SNPs achieving genome-wide significance in the discovery GWAS. Position is based on Human Genome Assembly build 37. Nearest gene or previously published gene names are included. Chr, chromosome; MAF, minor allele frequency; OR, odds ratio; 95% CI, confidence interval; *, rs113247976 represents the p.Pro986Leu variant in KIF5A (NM_004984.2).
### Table 2. Top ALS associations identified through RVB of FALS and control exome sequencing results

<table>
<thead>
<tr>
<th>Gene</th>
<th>FALS</th>
<th>Control</th>
<th>OR</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIF5A</td>
<td>6 (0.53%)</td>
<td>3 (0.02%)</td>
<td>32.07 (9.05-135.27)</td>
<td>5.55x10^-7</td>
</tr>
<tr>
<td>TBK1</td>
<td>8 (0.70%)</td>
<td>9 (0.05%)</td>
<td>15.11 (5.81-38.69)</td>
<td>5.58x10^-7</td>
</tr>
<tr>
<td>NEK1</td>
<td>12 (1.05%)</td>
<td>32 (0.16%)</td>
<td>6.64 (3.32-12.51)</td>
<td>1.68x10^-6</td>
</tr>
<tr>
<td>CALHM2</td>
<td>7 (0.62%)</td>
<td>9 (0.05%)</td>
<td>12.13 (4.47-31.79)</td>
<td>9.19x10^-6</td>
</tr>
<tr>
<td>COL14A1</td>
<td>8 (0.70%)</td>
<td>16 (0.08%)</td>
<td>8.04 (3.32-18.08)</td>
<td>2.72x10^-5</td>
</tr>
<tr>
<td>AK1</td>
<td>10 (0.88%)</td>
<td>34 (0.17%)</td>
<td>5.37 (2.55-10.41)</td>
<td>5.62x10^-5</td>
</tr>
<tr>
<td>ATRN</td>
<td>5 (0.44%)</td>
<td>9 (0.05%)</td>
<td>11.06 (3.57-31.02)</td>
<td>1.66x10^-4</td>
</tr>
<tr>
<td>VLDLR</td>
<td>5 (0.44%)</td>
<td>9 (0.05%)</td>
<td>10.87 (3.51-30.43)</td>
<td>1.79x10^-4</td>
</tr>
<tr>
<td>FUS</td>
<td>4 (0.35%)</td>
<td>4 (0.02%)</td>
<td>16.53 (4.25-64.33)</td>
<td>2.08x10^-4</td>
</tr>
<tr>
<td>ZMYND12</td>
<td>6 (0.53%)</td>
<td>12 (0.06%)</td>
<td>7.92 (2.86-19.96)</td>
<td>2.61x10^-4</td>
</tr>
<tr>
<td>Position</td>
<td>Variant</td>
<td>Exon</td>
<td>cDNA</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>------</td>
<td>------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Control Variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57,963,470</td>
<td>A&gt;G</td>
<td>11</td>
<td>c.1117+4A&gt;G</td>
<td>3' Splice Junction</td>
</tr>
<tr>
<td>57,966,423</td>
<td>C&gt;T</td>
<td>15</td>
<td>c.1630C&gt;T</td>
<td>p.Arg544*</td>
</tr>
<tr>
<td>57,976,884</td>
<td>G&gt;C</td>
<td>28</td>
<td>c.3021G&gt;C</td>
<td>5' Splice Junction</td>
</tr>
<tr>
<td>FALS Variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57,975,729</td>
<td>GA&gt;A</td>
<td>26</td>
<td>c.2987delA</td>
<td>p.Asp996fs</td>
</tr>
<tr>
<td>57,976,382</td>
<td>C&gt;T</td>
<td>27</td>
<td>c.2993-3C&gt;T</td>
<td>5' Splice Junction</td>
</tr>
<tr>
<td>57,976,385</td>
<td>GA&gt;G</td>
<td>27</td>
<td>c.2996delA</td>
<td>p.Asn999fs</td>
</tr>
<tr>
<td>57,976,411</td>
<td>A&gt;G</td>
<td>27</td>
<td>c.3019A&gt;G</td>
<td>p.Arg1007Gly</td>
</tr>
<tr>
<td>57,976,414</td>
<td>G&gt;A</td>
<td>27</td>
<td>c.3020+1G&gt;A</td>
<td>3' Splice Junction</td>
</tr>
<tr>
<td>57,976,415</td>
<td>T&gt;A</td>
<td>27</td>
<td>c.3020+2T&gt;A</td>
<td>3' Splice Junction</td>
</tr>
<tr>
<td>SALS Variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>57,957,481</td>
<td>G&gt;A</td>
<td>3</td>
<td>c.291+5G&gt;A</td>
<td>3' Splice Junction</td>
</tr>
<tr>
<td>57,975,731</td>
<td>CA&gt;C</td>
<td>26</td>
<td>c.2989delA</td>
<td>p.As997fs</td>
</tr>
<tr>
<td>57,976,384</td>
<td>G&gt;A</td>
<td>27</td>
<td>c.2993-1G&gt;A</td>
<td>5' Splice Junction</td>
</tr>
</tbody>
</table>

Table 3. Loss of function variants within KIF5A identified in probands. P, possible; Y, yes; N, no; M, male; F, female; L, limb onset; B, bulbar onset; n/a, not available or applicable. Note, ASSEDA does not predict exon skipping based on frameshifts or nonsense mutations (Tompson et al., 2007).
STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, John Landers (john.landers@umassmed.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study cohorts

GWAS cohort I. We undertook a GWAS of patients diagnosed with ALS (case cohort) and neurologically normal control individuals (control cohort). DNA was extracted from either whole blood or frozen brain tissue samples using standard procedures. All 12,663 patients included in the case cohort had been diagnosed with ALS according to the El Escorial criteria (Brooks, 1994) by a neurologist specializing in ALS, had onset of symptoms after age 18 years, and were of non-Hispanic white race/ethnicity. Both patients with familial ALS and patients with sporadic ALS were included in the analysis.

For the control cohort, we used genotype data obtained from (a) the database of Genotypes and Phenotypes (dbGaP) web repository (n = 44,017 US samples); (b) the HYPERGENES Project (n = 887 Italian samples) (Salvi et al., 2012); and (c) the Wellcome Trust Case Control Consortium (n = 5,663 British samples). An additional 2,112 US and Italian control samples were genotyped in the Laboratory of Neurogenetics, National Institute on Aging. The control cohort was matched to the case cohort for race and ethnicity, but not for age or sex. A detailed description of the cohorts is available in Table S1, S2.
Written consent was obtained from all individuals enrolled in this study, and the study was approved by the institutional review board approval of the National Institute on Aging (protocol number 03-AG-N329).

**GWAS cohort II.** Summary statistics from a recently published GWAS based on logistic regression analysis involving 12,577 cases and 23,475 controls were downloaded from the Project MinE Variant Browser. Additional details of the cohorts used in this study are available in van Rheenen *et al* (van Rheenen *et al.*, 2016).

**FALS discovery cohort.** A total of 1,463 FALS patients were included in the initial cohort (pre-QC). Patients were recruited at specialist clinics in Australia (n = 92), Belgium (n = 13), Canada (n = 34), Germany (n = 228), Ireland (n = 18), Israel (n = 26), Italy (n = 230), Netherlands (n = 50), Spain (n = 60), Turkey (n = 72), UK (n = 223), and USA (n = 417). All samples were exome sequenced except those from the Netherlands which were whole genome sequenced. Familial history was considered positive for ALS if the proband had at least one affected relative within three degrees of relatedness.

**Control discovery cohort.** Read level sequencing data were obtained from dbGAP and the European Genome-Phenome Archive (EGA) and are listed in Table S5.

**ALS WXS/WGS replication cohort.** Replication analyses included sequencing data for a further 9,046 ALS cases and 1,955 non-ALS controls that were not also represented in the FALS discovery set. These samples included 2,742 cases subjected to WXS by the ALS Sequencing Consortium, as described previously (Cirulli *et al.*, 2015); 719 cases subjected to WXS by the Laboratory of Neurogenetics,
National Institute on Aging; 307 cases and 296 controls subjected to WGS by the Laboratory of Neurogenetics, National Institute on Aging; 155 cases subjected to WGS by the CReATe Consortium; 1,017 cases subjected to WGS by the NYGC ALS Consortium, Genomic Translation for ALS Care (GTAC) Consortium and Answer ALS Foundation; and 4,100 cases and 1,659 controls subjected to WGS by the Project MinE Sequencing Consortium.

All samples included in the case cohort had been diagnosed with ALS according to the El Escorial criteria (Brooks, 1994) by a neurologist specializing in ALS. We received approval for this study from the institutional review boards of the participating centers, and written informed consent was obtained from all patients (consent for research).

METHOD DETAILS

Data generation and pre-processing

*Generation of SNP array callset.* The case cohort (n = 12,663 samples) and part of the control cohort (n = 2,112) were genotyped in the Laboratory of Neurogenetics, National Institute on Aging, using HumanOmniExpress BeadChips (version 1.0, Illumina Inc., San Diego, CA) according to the manufacturer's protocol. These SNP genotyping arrays assay 716,503 SNPs across the genome. Individual-level genotypes for these samples are available on the dbGaP web portal (accession number phs000101.v4.p1). The remainder of the control cohort had been previously genotyped on HumanOmni BeadChips (Illumina) as part of other GWAS efforts (see Table S2). Analyses were confined to the 595,692 autosomal SNPs that were common across the SNP genotyping arrays.
Generation of FALS case-control callset for exome-wide RVB discovery analysis. Exome sequencing of cases was performed as previously described (Kenna et al., 2016). Control exome sequences were generated as described under the relevant dbGAP and EGA project accessions. Sequence reads were aligned to human reference GRCh37 using BWA (Burrows-Wheeler Aligner) and processed according to recommended Genome Analysis Toolkit’s (GATK) best practices (https://software.broadinstitute.org/gatk/best-practices/). Joint variant detection and genotyping of all samples were performed using the GATK HaplotypeCaller. Variant quality control was performed using the GATK variant quality score recalibration method with default filters. A minimum variant quality by depth (QD) score of 2 was also imposed and all genotypes associated with genotype quality (GQ)< 20 were reset to missing. Variants were also excluded in the event of case or control call rates <70% (post genotype QC). Identified variants can be viewed through our web based ALS Variant Server (see link below).

Generation of ALS case-control callset for KIF5A replication analysis. Data for the KIF5A locus was extracted from all independently generated sequencing datasets and remapped to GRCh37. Variant calling was performed using the GATK haplotype caller as described above. In addition to the KIF5A locus, data was also extracted for a panel of 240,715 common variant sites and used to perform a single unified sample QC as described below.

Functional annotation of variants identified by WXS/WGS

Variant calls were assigned predicted functional consequences using snpEFF (Single Nucleotide Polymorphism Effect)(Cingolani et al., 2012), dbNSFP (A Database of Human Non-synonymous SNVs and Their Functional Predictions and Annotations)(Liu et al., 2013) and dbscSNV (database of splice
site consequences of Single Nucleotide Variants)(Jian et al., 2014), which is incorporated into dbNSFP. Variants were classified as "loss of function" (LOF) where the sequence change was predicted to encode a premature stop codon, a frameshift causing insertion-deletion or a splice site disrupting SNV. Variants were classified as potentially splice altering if assigned an "ada" or "rf" score >0.7 by dbscSNV. Splice variants of potential interest were further assessed for putative effects on exon skipping using a secondary algorithm - automated splice site and exon definition server (ASSEDA)(Tompson et al., 2007).

RT-PCR Analysis

Total RNA was prepared from lymphoblast lines using Trizol reagent. Reverse transcription using Applied Biosystems RNA to cDNA kit (# 4368814) was performed with 0.5 ug with RNAse inhibitor in a 20 ul reaction according to the manufacturer's protocol. PCR was carried out using New England Biolabs One Taq Hot Start DNA Polymerase (# M0481S), 2 ul RT reaction (representing 50 ng input RNA) and forward and reverse primer (0.15 uM each) in a 20 ul reaction volume. Amplification conditions were as follows: 94°C for 30 seconds, {94°C for 20 seconds, 58°C for 20 seconds, 68°C for 1 minute} x 35 cycles, followed by an extension stage of 68°C for 5 minutes and a 4°C hold. Amplification of both normal and mutant splice forms used primers F1 (CAGTGGAGCCACATCTTCTG) and R1 (TCTCTTGTGGAGAGGGAAA). Primers used for the specific amplification of the mutant splice form were F2 (CCACATGGACATGGAGTG), which spans exons 26 and 28, and R1.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses
Analysis of SNP array genotypes. Standard quality-control procedures were applied to our genotype data using PLINK software package (version 1.9) (Chang et al., 2015), and a summary of the workflow is shown in Figure S1. We excluded samples that demonstrated: call rates of less than 97.5%; non-European ancestry; abnormal F inbreeding coefficient; mismatch between phenotypic and genotypic gender; or, cryptic relatedness defined as identity-by-descent proportion of inheritance (pi_hat from PLINK) greater than 0.125. Samples in common between our study and van Rheenen's study were identified using the checksum program id_geno_checksum and were removed from our analyses. We excluded palindromic SNPs, as well as SNPs with: call rates less than 95% in the US and Italian cohorts or less than 99% in the UK, French and Belgium cohorts; minor allele frequency less than 0.05 in the control cohorts; Hardy-Weinberg equilibrium P less than 10^{-7} in the US and Italian control cohorts and less than 10^{-5} in the UK, French and Belgium cohorts; missingness by case-control status P less than 10^{-5}; or SNPs associated between the UK and French control cohorts with P less than 5.0x10^{-8}. After quality control, 8,229 case and 36,329 control samples were included in the analysis, and 436,746 SNPs were available for imputation in the USA and Italy cohorts, and 420,131 SNPs were available in the UK, French and Belgium cohorts.

Estimation of the haplotypes was performed with SHAPEIT (version 2.r790) (Delaneau et al., 2013). Imputation was performed for individual batches based on ethnicity using the 1000 Genomes Project dataset (phase 3, version 5a, release 2013-05-02, 1000genomes.org) as reference and using Minimac3 software (version 1.0.11) (Das et al., 2016) with default settings. After imputation, principal components were calculated using PLINK software after removing known hypervariable regions and the 1 MB surrounding the C9orf72 region. After analysis of the Scree plots, 2 to 4 principal components were
retained per cohort as covariates in the association analyses to compensate for any residual population stratification.

Logistic regression was performed per batch using mach2dat software (version 1.0.24) (Marchini and Howie, 2010) incorporating 2 to 4 principal components, age and gender as covariates, with dosage of imputed SNPs selected based on a Minimac3 R² value of imputation accuracy greater than 0.3. SNPs with an absolute beta coefficient value above 5 or with a minor allele frequency less than 0.01 were excluded from meta-analysis. Meta-analysis was then performed combining the association results of the 13 batches of our individual-level studies with van Rheenen's study summary statistics using METAL software (version 2011-03-25) (Willer et al., 2010) under an inverse-weighted, fixed effect model. A threshold $P$ of $5.0 \times 10^{-8}$ was set for genome-wide significance after Bonferroni correction for multiple testing in the GWAS (Pe'er et al., 2008).

The programming code used to analyze these data is freely available on GitHub (see link below), and GWAS summary statistics results for all tested SNPs are available from (link to be supplied at publication).

*Analysis of WXS/WGS genotypes.* For both the discovery and replication phases, samples were excluded from the study in the event of failing to meet standard genotype call rate, heterozygosities, duplication, relatedness or population stratification filters as summarized in Table S6. Each of these filters was performed using a set of autosomal markers meeting all of the following criteria: call rate $> 0.95$, minor allele frequency (MAF) $> 0.01$, $P > 0.001$ for deviation from Hardy-Weinberg equilibrium, linkage disequilibrium pruning ($R^2 < 0.5$, window size = 50, step = 5). Filtering of autosomal markers, sample
call rate assessments and sample heterozygosity assessments were performed using PLINK software. Study duplicates and sample relatedness within the WXS/WGS cohorts was identified using KING software (Manichaikul et al., 2010). Study duplicates between WXS/WGS cohorts and GWAS datasets were identified using the checksum program id_geno_checksum. LASER was used to generate PCA coordinates for samples from the Human Genome Diversity Panel (HGDP). Samples from the FALS discovery cohort were then mapped to this reference co-ordinate space. The discovery cohort was restricted to cases and controls occurring within 3 standard deviations of the mean for European HGDP samples along principal components 1-4.

RVB analyses were performed by penalized logistic regression of case-control status with respect to number of minor alleles observed per sample per gene with and MAF <0.001. Analyses were only performed where the dataset contained more than 3 variant allele occurrences. Replication analyses of rs113247976 were performed using the same logistic regression protocol as used for RVB analyses. All analyses were conditioned on the first 4 eigenvectors generated by principal components analysis of common variant profiles. Genomic inflation factors were calculated using genome-wide association analysis for quantitative, binary and time-till-event traits using GenABEL software. Candidate associations were tested for signs of call-rate or subcohort biases as outlined in Figures S6, S7. Meta-analysis of rs113247976 association results between sequencing and GWAS was performed using METAL. Unless otherwise indicated, all statistical analyses were performed using R (version 3.2.0).

Control-control analyses. To identify genes potentially subject to confounding biases in FALS RVB analyses and to assess the potential impact of batch effects with non-ALS-related data, population or phenotypic stratifiers, the control sample cohort was divided into 28 pseudo case-control groups based
on the sequencing center or associated dbGaP / EGA project (Table S5). Genes shown in gray achieve for possible confounder association. Loci achieving a minimu $P < 1 \times 10^{-3}$ were deemed as displaying possible association with non-ALS related batch effects.

**DATA AND SOFTWARE AVAILABILITY**

**Datasets**
The programming code used to analyze the GWAS data including the imputation with SHAPEIT and Minimac3, individual-based association analysis using Mach2dat and a meta-analysis using METAL is freely available on GitHub: https://github.com/AudeDN/ALS_GWAS_1000G_mach2dat_2017). GWAS summary statistics results for all tested SNPs and identified SNVs from our 1,138 FALS cohort used for the RVB analysis can be viewed through our web based ALS Variant Server (http://als.umassmed.edu). For each variant, information on over 50 annotation fields and the results can be downloaded directly into Excel.

**Data Resources and Databases**

1000 Genomes Project dataset: http://www.internationalgenome.org
dbNSFP: https://sites.google.com/site/jpopgen/dbNSFP
dbscSNV: incorporated into dbNSFP (see previous link).
European Genome-phenome Archive (EGA): https://ega-archive.org
Human Genome Diversity Panel (HGDP): http://www.hagsc.org/hgdp/
HYPERGENES Project: http://www.hypergenes.eu
Project MinE Variant Browser: http://databrowser.projectmine.com
snpEFF: http://snpeff.sourceforge.net/SnpEff.html

Wellcome Trust Case Control Consortium: www.wtccc.org.uk

Software

ASSEDA: http://www.cytognomix.com/?post_type=duka&p=2670

BWA: http://bio-bwa.sourceforge.net

GenABEL: http://www.genabel.org

GATK: https://software.broadinstitute.org/gatk/

id_geno_checksum: https://personal.broadinstitute.org/sripke/share_links/checksums_download/

KING: http://people.virginia.edu/~wc9c/KING/

LASER: http://csg.sph.umich.edu/chaolong/LASER/

Mach2dat: https://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output

METAL: http://csg.sph.umich.edu/abecasis/metal/index.html

Minimac3: https://genome.sph.umich.edu/wiki/Minimac3

PLINK: http://zzz.bwh.harvard.edu/plink/

R: https://www.r-project.org

SHAPEIT: https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html
SUPPLEMENTAL ITEM TITLES

Figure S1. Related to Figure 1; Workflow showing the quality control procedures applied to the present study.

Figure S2. Related to Figure 1; Multi-dimensional scaling plot of the 44,558 genotyped samples included in analysis compared to the HapMap populations.

Figure S3. Related to Figure 1; Quartile-Quartile plot of $P$-values from the meta-analysis based on logistic regression analysis.

Figure S4. Related to Figure 3; Principal components analysis of samples included in the RVB analysis compared to the Human Diversity Panel.

Figure S5. Related to Figure 3; Quartile-Quartile plot of $P$ values from the gene-based rare variant burden analysis of exome data.

Figure S6. Related to Figure 3; Control-control analyses.

Figure S7. Related to Figure 3; Plot of variant call rates across the KIF5A protein-coding region in FALS versus controls analyzed by RVB testing.

Figure S8. Related to Figure 2; Principal components analysis of samples included in KIF5A replication cohort.

Table S1. Related to Figure 1; Demographics and baseline characteristics of patients diagnosed with ALS and control individuals included in the GWAS analysis.

Table S2. Related to Figure 1; DbGaP studies contributing to the GWAS analysis.

Table S3. Related to Figure 1; SNPs achieving genome-wide significance in the GWAS analysis.

Table S4. Related to Figure 1; Suggestive SNPs with $P$ values less than $5.0 \times 10^{-7}$ in the GWAS analyses.

Table S5. Related to Figure 3; DbGaP/EGA studies contributing to the RVB analysis.

Table S6. Related to Figure 2, 3; Quality control filtering of the FALS discovery and KIF5A replication cohorts.
<table>
<thead>
<tr>
<th>Cohort</th>
<th>Cases</th>
<th>Controls</th>
<th>Case MAF</th>
<th>Control MAF</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italy</td>
<td>2,853</td>
<td>2,143</td>
<td>1.28%</td>
<td>0.82%</td>
<td>1.45 (0.90 - 2.34)</td>
<td></td>
</tr>
<tr>
<td>United Kingdom</td>
<td>449</td>
<td>226</td>
<td>3.56%</td>
<td>1.33%</td>
<td>2.07 (0.85 - 5.07)</td>
<td></td>
</tr>
<tr>
<td>Belgium / France</td>
<td>1,150</td>
<td>595</td>
<td>1.91%</td>
<td>1.26%</td>
<td>1.42 (0.75 - 2.68)</td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>3,777</td>
<td>33,365</td>
<td>2.01%</td>
<td>1.46%</td>
<td>1.45 (1.19 - 1.75)</td>
<td></td>
</tr>
<tr>
<td><strong>Present Study</strong></td>
<td><strong>8,229</strong></td>
<td><strong>36,329</strong></td>
<td><strong>1.83%</strong></td>
<td><strong>1.42%</strong></td>
<td><strong>1.46 (1.24 - 1.73)</strong></td>
<td><strong>9.23 x 10^-6</strong></td>
</tr>
<tr>
<td>Van Rheenen Study</td>
<td>12,577</td>
<td>23,475</td>
<td>2.23%</td>
<td>1.75%</td>
<td>1.33 (1.17 - 1.52)</td>
<td>1.13 x 10^-5</td>
</tr>
<tr>
<td>Combined Discovery Set</td>
<td>20,806</td>
<td>59,804</td>
<td>2.07%</td>
<td>1.55%</td>
<td>1.38 (1.25 - 1.53)</td>
<td>6.43 x 10^-10</td>
</tr>
<tr>
<td>Replication</td>
<td>4,159</td>
<td>18,650</td>
<td>1.78%</td>
<td>1.32%</td>
<td>1.42 (1.17 - 1.70)</td>
<td>3.82 x 10^-4</td>
</tr>
<tr>
<td><strong>Joint</strong></td>
<td><strong>24,965</strong></td>
<td><strong>78,454</strong></td>
<td><strong>2.02%</strong></td>
<td><strong>1.49%</strong></td>
<td><strong>1.39 (1.33 - 1.45)</strong></td>
<td><strong>7.09 x 10^-13</strong></td>
</tr>
</tbody>
</table>
A. Consensus Splice Sequence:

B. Normal Splice RNA:

C. Normal Splice mRNA:

D. c.3020+2T>A

D. c.3020+1G>A
SUPPLEMENTAL MATERIAL

Author Affiliations

1. Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institutes on Aging, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892, USA
2. These authors contributed equally to this work
3. Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts, USA.
4. Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
5. Ronald M. Loeb Center for Alzheimer's Disease, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
6. Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy.
7. Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center - Università degli Studi di Milano, Milan 20122 Italy.
8. Molecular Genetics Section, Laboratory of Neurogenetics, National Institutes on Aging, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892, USA
9. Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA
10. Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.
11. Data Tecnica International, Glen Echo, MD, USA
12. Department of Neurology Brain Centre, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, GA Utrecht, The Netherlands
13. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands.
14. Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
15. United Kingdom Dementia Research Institute Centre, Maurice Wohl Clinical Neuroscience Institute, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, 125 Coldharbour Lane, Camberwell, SE5 9NU London, U.K.
16. Institute of Genomic Medicine, Catholic University, Roma, Italy
17. Sobell Department of Motor Neuroscience and Movement Disorders, University College London, Institute of Neurology, London, UK
18. Genetics, Genetics and Pharmacogenomics, MRL, Merck & Co., Inc., Boston, MA 02115, USA
19. ALS Center, Salvatore Maugeri Foundation, Milan, Italy
20. ‘Rita Levi Montalcini’ Department of Neuroscience, University of Turin, Turin, Italy
21. ALS Center, Department of Neurology, Azienda Ospedaliera Universitaria Maggiore della Carità, Novara, Italy
22. Department of Neurology, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy
23. Department of Neuroscience, St. Agostino Estense Hospital, Azienda Ospedaliero Universitaria di Modena, Modena, Italy
24. Department of Neurosciences, Ophthalmology, Genetics, Rehabilitation, Maternal and Child Health, Ospedale Policlinico San Martino, Genoa, Italy
25. Department of Medical, Surgical and Neurological Sciences, University of Siena, Siena, Italy
26. ALS Center, Salvatore Maugeri Foundation, IRCCS, Mistretta, Messina, Italy
27. ALS Clinical Research Center, University of Palermo, Palermo, Italy
28. Institute of Neurological Sciences, National Research Council, Mangone, Cosenza, Italy
29. Department of Neurology, Azienda Universitario Ospedaliera di Cagliari and University of Cagliari, Cagliari, Italy
30. Department of Clinical and Experimental Medicine, University of Messina and Nemo Sud Clinical Center for Neuromuscular Diseases, Aurora Foundation, Messina, Italy
31. Department of Basic Medical Sciences, Neurosciences and Sense Organs, University of Bari, Italy
32. Department of Medical, Surgical, Neurological, Metabolic and Aging Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy
33. Il Bene Center for Immunological and Rare Neurological Diseases at Bellaria Hospital, IRCCS, Istituto delle Scienze Neurologiche, Bologna, Italy
34. Neurological Clinic, Marche Polytechnic University, Ancona, Italy
35. Department of Health Sciences, University of Eastern Piedmont, Novara, Italy
36. Department of Neurology, University of Chieti, Chieti, Italy
37. Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, National Institutes of Health, Baltimore, MD, 21224, USA
38. Department of Neurology, Columbia University, New York, NY 10032 USA.
39. Institute for Genomic Medicine, Columbia University, New York, NY 10032 USA.
40. Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305 USA.
41. Bioverativ, 225 2nd Ave, Waltham, MA 02145
42. HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806 USA.
43. Center for Genomics of Neurodegenerative Diseases (CGND), New York Genome Center, New York, NY
44. Computational Biology, New York Genome Center, New York, NY
45. Gladstone Institute of Neurological Disease, San Francisco, CA, USA
46. Departments of Neurology and Physiology, University of California, San Francisco, CA, USA
47. Gladstone Institutes, San Francisco CA 94158
48. Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, 02139, USA
49. Broad Institute, 415 Main St, Cambridge, Massachusetts, 02142, USA
50. Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA.
51. Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA.
52. Department of Neurobiology and Behavior, Institute of Memory Impairment and Neurological Disorders, University of California, Irvine, CA 92697, USA
53. Department of Psychiatry and Human Behavior, Institute of Memory Impairment and Neurological Disorders, University of California, Irvine, CA 92697, USA.
54. The Heart Institute and Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA
55. Harvard Medical School, Department of Neurology, Massachusetts General Hospital (MGH), Boston, Massachusetts, USA
56. Neurological Clinical Research Institute (NCRI), Massachusetts General Hospital, Boston, Massachusetts, USA
57. Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA.
58. Department of Neurology, The Ohio State University Wexner Medical Center, Columbus, OH, United States
59. Department of Neurology, The Ohio State University Wexner Medical Center, Columbus, OH, United States.
60. Harvard Medical School, Department of Neurology, Massachusetts General Hospital (MGH), Boston, Massachusetts, USA.
61. Neurological Clinical Research Institute (NCRI), Massachusetts General Hospital, Boston, Massachusetts, USA.
62. Department of Neurology, Johns Hopkins University, Baltimore, MD 21287, USA
63. Department of Neurology, University of Miami
64. Howard Hughes Medical Institute, Chevy Chase, MD 20815
65. Department of Cell and Molecular Biology, St. Jude Children’s Research Hospital, Memphis, TN 38105
66. Department of Computational Biology, St. Jude Children’s Research Hospital, Memphis, TN 38105
67. 3rd Neurology Unit, Motor Neuron Diseases Center, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan, Italy and Department of Biomedical and Clinical Sciences "Luigi Sacco", University of Milan, Italy
68. Department of Neurology, IRCCS Istituto Auxologico Italiano, Milan, Italy
69. Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, King’s College London, London SE5 9RS, UK
70. Neurology Unit, IRCCS Foundation Ca’ Granda Ospedale Maggiore Policlinico, Milan, Italy.
71. Department of Neurosciences, University of Padova, Padova, Italy
72. Experimental Neurobiology Laboratory, ‘C. Mondino’ National Institute of Neurology Foundation, IRCCS, Pavia, Italy
73. ALS Center, CHU Bretonneau, Tours university, Tours, France
74. Department of Neurology, Helsinki University Central Hospital and Molecular Neurology Programme, Biomedicum, University of Helsinki, Helsinki, FIN-02900, Finland
75. Department of Pathology, Medicum, University of Helsinki, and HUSLAB, FIN-00014 Helsinki, Finland
76. Department of Neurology, Helsinki University Central Hospital and Molecular Neurology Programme, Biomedicum, University of Helsinki, Helsinki, FIN-02900, Finland
77. Dept of Neurology, Salford Royal Foundation Trust, Stott Lane, Salford, M6 8HD
78. Faculty of Human and Medical Sciences, University of Manchester, Manchester M13 9PT, UK
79. Department of Clinical Neuroscience, Institute of Neurology, University College London, London NW3 2PG, UK
80. Department of Molecular Neuroscience and Reta Lila Weston Laboratories, Institute of Neurology, University College London, Queen Square House, London WC1N 3BG, UK
81. Centre for Neuroscience and Trauma, Blizard Institute, Queen Mary University of London, North-East London and Essex Regional Motor Neuron Disease Care Centre, London, E1 2AT, UK
82. Genomics Technology Group, Laboratory of Neurogenetics, National Institutes on Aging, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892, USA
83. Department of Neurology, Emory University, Atlanta, GA 30322, USA.
84. Division of Brain Sciences, Department of Medicine, Imperial College London, W120NN, UK
85. Department of Genome analysis and Neurogenetics, Academic Medical Centre, Amsterdam, The Netherlands
86. Department of Neurogenetics and Neurology, Academic Medical Centre, Amsterdam, The Netherlands
87. Unidad de ELA, Instituto de Investigación Hospital Gregorio Marañón de Madrid, SERMAS, Spain.
88. Computational Biology Core, Laboratory of Neurogenetics, National Institutes on Aging, National Institutes of Health, Porter Neuroscience Research Center, Bethesda, MD 20892, USA
89. Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA
90. Motor Neuron Disorders Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA
91. Center for Geriatric Medicine, Department of Geriatrics, Neurosciences and Orthopedics, Catholic University of Sacred Heart, Rome 00168, Italy
92. Division of Neurology, Barrow Neurological Institute, Phoenix, Arizona, USA
93. Departments of Neurology, University of Utah School of Medicine, Salt Lake City, Utah, USA
94. Department of Neuroscience, University of California San Diego, La Jolla, CA, USA
95. Mount Sinai Beth Israel Hospital, Mount Sinai School of Medicine, NYC NY
96. Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK
97. Department of Neurology, Neuromuscular Center, Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA
98. Discipline of Pathology, Brain and Mind Centre, The University of Sydney, 94 Mallett St, Camperdown, NSW 2050 Australia
99. Department of Biochemistry, Penn State College of Medicine, Hershey, PA, USA
100. Department of Pathology, Penn State College of Medicine, Hershey, PA, USA
101. Research and Development Service, Veterans Affairs Boston Healthcare System, Boston, MA
102. Department of Neurology, Program in Behavioral Neuroscience, Boston University School of Medicine, Boston, MA
103. Neurology Service, VA Boston Healthcare System and Boston University Alzheimer's Disease Center, Boston MA 02130
104. Departments of Pathology and Neurology, Johns Hopkins University School of Medicine
105. Sorbonne Universités, UPMC Univ Paris 06, Inserm, CNRS, Institut du cerveau et la moelle (ICM), Assistance Publique-Hôpitaux de Paris (AP-HP) - Hôpital Pitié-Salpêtrière, Paris, France
106. Service de biochimie, CHU de Nîmes, Nîmes, France
107. Department of Neurology, Drexel University College of Medicine, Philadelphia, PA
108. Department of Neurology, Lewis Katz School of Medicine, Temple University, Philadelphia, PA
109. Epidemiology Branch, National Institute of Environmental Health Sciences, NC 27709, USA.
110. KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology, Leuven, Belgium
111. VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium
112. Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, California, USA
113. Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
114. Institute of Human Genetics, Technische Universität München, Munich, Germany.
115. Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, UK.
116. Institute of Physiology, Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon, Portugal.
117. Department of Neurosciences, Hospital de Santa Maria-CHLN, Lisbon, Portugal.
118. SURFsara, Amsterdam, the Netherlands.
119. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands.
120. KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium.
121. VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium.
122. Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland.
123. Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, St. Gallen, Switzerland.
124. Department of Neurology, Mayo Clinic Florida, Jacksonville, Florida 32224, USA.
125. Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, USA.
126. Faculty of Medicine, University of Southampton, Southampton, UK.
127. Suna and Inan Kirac Foundation, Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey.
128. ALS Unit/Neurology, Hospital San Rafael, Madrid, Spain.
129. Department of Neurology Tel-Aviv Sourasky Medical Centre, Israel.
130. Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK.
133. Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Republic of Ireland.
134. Centre for MND Research, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW 2109, Australia.
135. ANZAC Research Institute, Concord Hospital, University of Sydney, Sydney, NSW 2139, Australia.
136. Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada.
137. Unidad de ELA, Instituto de Investigación Hospital 12 de Octubre de Madrid, SERMAS, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER U-723), Madrid, Spain.
138. Tanz Centre for Research of Neurodegenerative Diseases, Division of Neurology, Department of Medicine, University of Toronto, Toronto, ON, M5S 3H2, Canada.
139. Division of Neurology, Department of Internal Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON, M4N 3M5, Canada.
140. Department of Neurology, Penn State Hershey Medical Center, Hershey, PA, USA.
141. Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK.
142. Department of Neurology, University of Michigan, Ann Arbor, Michigan, USA.
143. Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan 20133, Italy.
144. Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan 20149, Italy.
145. Department of Neurology, University Hospital KU Leuven (University of Leuven) and Department of Neurosciences, KU Leuven, B-3000 Leuven, Belgium
146. Centro Clinico NeMO, Institute of Neurology, Catholic University, Largo F. Vito 1, 00168 Rome, Italy
147. NEuroMuscular Omnicenter (NEMO), Serena Onlus Foundation, Milan, Italy
148. Neurology Department, Ulm University, Albert-Einstein-Allee 11, 89081 Ulm, Germany
149. Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå SE-90185
150. ALS center, CHU Gui de Chauliac, University of Montpellier, Montpellier, France
151. Department of Neurology, Gui-de-Chauliac Hospital, Montpellier, France
152. Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
153. Department of Neurology, Columbia University, New York, NY, USA
154. Department of Neurology, Emory University School of Medicine, Atlanta, GA
155. Genetics, Genetics and Pharmacogenomics, MRL, Merck & Co., Inc., West Point, PA 19486, USA
156. Neuroscience Institute of Torino, Turin 10124, Italy
157. Lead Contact

* Correspondence: john.landers@umassmed.edu; bryan.traynor@nih.gov

**Consortium Members**

**ALS Sequencing Consortium**


1. Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, NC 27708 USA.
2. Stanley Appel Department of Neurology, Houston Methodist, Houston, TX USA.
3. Department of Neurology, Cedars Sinai Medical Center, Los Angeles, CA 90048 USA.
4. Duke ALS Clinic and Durham VA Medical Center, Duke University School of Medicine, Durham, NC 27708 USA.
5. HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806 USA.
6. Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01655 USA.
7. Biogen Idec, Cambridge, MA 02142 USA.
8. Department of Genetics, Stanford University School of Medicine, Stanford, CA 94305 USA.
9. Department of Pediatrics and Medicine, College of Physicians and Surgeons, Columbia University, New York NY 10032
10. Human Longevity, INC, San Diego, CA 92121 USA.
11. Department of Neurology and Neurosurgery, McGill University, Montreal, H3A 2B4 Canada
12. Department of Neurology, University of Utah School of Medicine, Salt Lake City, UT 84112 USA.
13. Department of Neurology, Emory University, Atlanta, GA 30322 USA.
14. Institute for Genomic Medicine, Columbia University, New York, NY 10032 USA.
15. Duke University School of Medicine, Duke University School of Medicine, Durham, NC 27708 USA.
16. Department of Neurology/Institute for Genomic Medicine, Columbia University, New York, NY 11032 USA.
17. Bioverativ, 225 2nd Ave, Waltham, MA 02145
18. Biogen Idec and Harvard Medical School, Cambridge, MA 02142 USA.
19. Department of Biochemistry & Molecular Biophysics, Columbia University, New York, NY 10027 USA.
20. Department of Neurology, Washington University School of Medicine, St. Louis, MO 63110 USA.
21. Department of Neurosciences, University of California, San Diego, La Jolla CA 92039
22. Department of Neurology/Eleanor and Lou Gehrig ALS Center, Columbia University, New York, NY 10032 USA.
23. Neurogenetics DNA Diagnostic Laboratory, Center for Human Genetics Research Massachusetts General Hospital, Boston, MA 02114 USA.

**Answer ALS Foundation**

Julia Kaye¹, Steven Finkbeiner¹,², Stacia Wyman³, Alexander Lenail⁴, Leandro Lima¹, Ernest Fraenkel⁴,⁵, Jeffrey D Rothstein⁶, Clive N Svendsen⁷,⁸, Leslie M Thompson⁹,¹⁰, Jenny Van Eyk¹¹, Nicholas J Maragakis⁶, James D Berry¹²,¹³, Jonathan D Glass¹⁴, Timothy M Miller¹⁵, Stephen J Kolb¹⁶,¹⁷, Robert H Baloh¹⁸, Merit Cudkowicz¹⁹,²⁰, Emily Baxi⁶

1. Gladstone Institute of Neurological Disease, San Francisco, CA, USA
2. Departments of Neurology and Physiology, University of California, San Francisco, CA, USA
3. Gladstone Institutes, San Francisco CA 94158
4. Department of Biological Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, 02139, USA
5. Broad Institute, 415 Main St, Cambridge, Massachusetts, 02142, USA
6. Department of Neurology, Johns Hopkins University, Baltimore, MD 21287, USA
7. Board of Governors Regenerative Medicine Institute, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA
8. Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, 90048, USA
9. Department of Neurobiology and Behavior, Institute of Memory Impairment and Neurological Disorders, University of California, Irvine, CA 92697, USA
10. Department of Psychiatry and Human Behavior, Institute of Memory Impairment and Neurological Disorders, University of California, Irvine, CA 92697, USA.
11. The Heart Institute and Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA
12. Harvard Medical School, Department of Neurology, Massachusetts General Hospital (MGH), Boston, Massachusetts, USA
13. Neurological Clinical Research Institute (NCRI), Massachusetts General Hospital, Boston, Massachusetts, USA
14. Department of Neurology, Emory University School of Medicine, Atlanta, GA
15. Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
16. Department of Biological Chemistry & Pharmacology, The Ohio State University Wexner Medical Center, Columbus, OH, United States
17. Department of Neurology, The Ohio State University Wexner Medical Center, Columbus, OH, United States
18. Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, California, USA
19. Harvard Medical School, Department of Neurology, Massachusetts General Hospital (MGH), Boston, Massachusetts, USA.
20. Neurological Clinical Research Institute (NCRI), Massachusetts General Hospital, Boston, Massachusetts, USA

Clinical Research in ALS and Related Disorders for Therapeutic Development (CReATE) Consortium
Michael Benatar¹, J. Paul Taylor²³, Gang Wu¹, Evadnie Rampersaud⁴, Joanne Wu¹, Rosa Rademakers⁵, Stephan Züchner⁶, Rebecca Schule⁷, Jacob McCauley⁶, Sumaira Hussain¹, Anne Cooley¹, Marielle Wallace¹, Christine Clayman¹, Richard Barohn⁸, Jeffrey Statland⁹, John Ravits⁹, Andrea Swenson¹⁰, Carlayne Jackson¹¹, Jaya Trivedi¹², Shaida Khan¹², Jonathan Katz¹³, Liberty Jenkins¹⁵, Ted Burns¹⁴, Kelly Gwathmey¹⁴, James Caress¹⁵, Corey McMillan¹⁶, Lauren Elman¹⁶, Erik Pioro¹⁷, Jeannine Heckmann¹⁸, Yuen So¹⁹, David Walk²⁰, Samuel Maiser²¹, Jinghui Zhang⁴

1. Department of Neurology, University of Miami
2. Howard Hughes Medical Institute
3. Department of Cell and Molecular Biology, St. Jude Children’s Research Hospital
4. Department of Computational Biology, St. Jude Children’s Research Hospital
5. Department of Neuroscience, Mayo Clinic Jacksonville
6. Department of Human Genetics, University of Miami
7. University of Tubingen, Hertie Institute for Clinical Brain Research
8. Department of Neurology, University of Kansas Medical Center
9. Department of Neurosciences, University of California San Diego
10. Department of Neurology, University of Iowa
11. Department of Neurology, University of Texas Health Sciences Center San Antonio
12. Department of Neurology & Neurotherapeutics, University of Texas Southwestern Medical Center
13. Department of Neurology, California Pacific Medical Center
14. Department of Neurology, University of Virginia
15. Department of Neurology, Wake Forest University
16. Department of Neurology, University of Pennsylvania
17. Neuromuscular Center, Cleveland Clinic Ohio
18. Department of Neurology, University of Cape Town
19. Department of Neurology, Stanford University
20. Department of Neurology, University of Minnesota
21. Hennepin County Medical Center
French ALS Consortium
William Camu¹, Kevin Mouzat², Serge Lumbruso², Philippe Corcia³, Vincent Meininger³, Gérard Besson⁴, Emmeline Lagrange⁵, Pierre Clavelou⁵, Nathalie Guy⁶, Philippe Couratier⁷, Patrick Vourch⁸, Véronique Danel⁹, Emilien Bernard¹⁰, Gwendal Lemasson¹¹

1. ALS center, CHU Gui de Chauliac, University of Montpellier, Montpellier, France
2. Service de biochimie, CHU de Nîmes, Nîmes, France
3. ALS Center, CHU Bretonneau, Tours university, Tours, France
4. Service de Neurologie, Hôpital Michalon, La Tronche, France
5. Service d'explorations du système nerveux, Hôpital Michalon, La Tronche, France
6. ALS center, CHU Gabriel Montpied, Clermont-Ferrand, France
7. ALS center, CHU Dupuytren, Limoges, France
8. Département de Biochimie, CHU Bretonneau, Tours, France
9. ALS center, CHRU, Lille, France,
10. ALS center, CHU Wertheimer, Lyon, France
11. ALS center, CHU Pellegrin, Bordeaux, France

Genomic Translation for ALS Care (GTAC) Consortium
Matthew B. Harms¹, David B. Goldstein², Neil A. Shneider³, Stephen Goutman⁴, Zachary Simmons⁵, Timothy M. Miller⁶, Siddharthan Chandran⁷, Suvankar Pal⁷, George Manousakis⁸, Stanley H. Appel⁹, Ericka Simpson⁹, Leo Wang¹⁰, Robert H Baloh¹¹, Summer Gibson¹², Richard Bedlack¹³, David Lacomis¹⁴, Dhruv Sareen¹⁵, Alexander Sherman¹⁶, Lucie Bruijn¹⁷, Michelle Penny¹⁸

1. Department of Neurology/Institute of Genomic Medicine/Elanor and Lou Gehrig ALS Center, Columbia University, New York, New York 10032, USA
2. Institute for Genomic Medicine, Columbia University, New York, NY 10032, USA
3. Department of Neurology/Elanor and Lou Gehrig ALS Center, Columbia University, New York, NY 10032, USA
4. Department of Neurology, University of Michigan, Ann Arbor, Michigan 48109, USA
5. Department of Neurology, The Pennsylvania State University College of Medicine, Hershey, Hershey, PA 17033, USA
6. Department of Neurology, Washington University in St. Louis, St. Louis, MO, USA
7. Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, Scotland, UK
8. Department of Neurology, University of Minnesota, Minneapolis, MN, USA
9. Stanley Appel Department of Neurology, Houston Methodist, Houston, TX, USA
10. Department of Neurology, University of Washington, Seattle, WA, USA
11. Department of Neurology, Cedars Sinai Medical Center, Los Angeles, CA, USA
12. Department of Neurology, University of Utah, Salt Lake City, UT, USA
13. Duke ALS Clinic and Durham VA Medical Center, Duke University School of Medicine, Durham, NC 27708, USA
14. Department of Neurology and Pathology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA
15. Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA
16. Neurological Clinical Research Institute, Massachusetts General Hospital, Boston, MA, USA
ITALSGEN Consortium
Francesco O. Logullo1, Isabella Simone2, Giancarlo Logroscino2,3, Fabrizio Salvi4, Ilaria Bartolomei4, Giuseppe Borghero5, Maria Rita Murr5, Emanuela Costantino5, Carla Pani5, Roberta Puddu5, Carla Caredda4, Valeria Piras5, Stefania Tranquilli5, Stefania Cuccu6, Daniela Corongiu5, Maurizio Melis5, Antonio Milia5, Francesco Marrosu5, Maria Giovanna Marrosu5, Gianluca Floris5, Antonino Cannas5, Stefania Tranquilli5, Margherita Capasso6, Claudia Caponnetto7, Gianluigi Mancardi7, Paola Origone7, Paola Mandich8, Francesca L. Conforti8, Sebastiano Cavallaro8, Gabriele Morà8, Kalliopi Marinou9, Riccardo Sideri9, Silvana Penco10,13, Lorena Mosca10, Christian Lunetta11, Giuseppe Lauria Pinter12, Massimo Corbo13, Nilo Riva14, Paola Carrera14, Paolo Volanti15, Jessica Mandrioli16, Nicola Fini16, Antonio Fasano16, Lucio Tremolizzo17, Alessandro Arosio17, Carlo Ferrarese17, Francesca Trojsi18, Gioacchino Tedeschi19, Maria Rosaria Monsurrò19, Giovanni Piccirillo19, Cinzia Femiano19, Anna Ticca20, Enzo Ortu21, Vincenzo La Bella22, Rossella Spataro22, Tiziana Colletti22, Mario Sabatelli23, Marcella Zollino23, Amelia Conte23, Marco Luigetti23, Serena Lattante23, Giuseppe Marangi23, Marialuisa Santarelli24, Antonio Petrucci25, Maura Pugliatti26, Angelo Pirisi26, Leslie D. Parish26, Patrizia Occhini26, Fabio Giannini27, Stefania Battistini27, Claudia Ricci27, Michele Benigni27, Tea B. Cau28, Daniela Loi28, Andrea Calvo29, Cristina Moglia29, Maura Brunetti30, Marco Barberis30, Gabriella Restagno30, Federico Casale30, Giuseppe Marrali30, Giuseppe Fuda30, Irene Ossola30, Stefania Cammarosano30, Antonio Canosa30, Antonio Ilardi30, Umberto Manera30, Maurizio Grassano30, Raffaella Tanel31, Fabrizio Pisano32

1. Neurological Clinic, Marche Polytechnic University, Ancona, Italy
2. Department of Basic Medical Sciences, Neurosciences and Sense Organs, University of Bari, Bari, Italy
3. Department of Clinical and Research Neurology, “Pia Fondazione Cardinal G. Panico” Hospital, Tricase (LE), Bari, Italy
4. Center for Diagnosis and Cure of Rare Diseases, Department of Neurology, IRCCS Institute of Neurological Sciences, Bologna, Italy
5. Department of Neurology, Azienda Universitaria Ospedaliera di Cagliari and University of Cagliari, Cagliari, Italy
6. Department of Neurology, University of Chieti, Chieti, Italy
7. Department of Neurosciences, Ophthalmology, Genetics, Rehabilitation, Maternal and Child Health, Ospedale Policlinico San Martino, Genoa, Italy
8. Institute of Neurological Sciences, National Research Council, Mangone, Cosenza, Italy
9. Department of Neurological Rehabilitation, Fondazione Salvatore Maugeri, IRCCS, Istituto Scientifico di Milano, Milano, Italy
10. Department of Laboratory Medicine, Medical Genetics, Niguarda Ca’ Granda Hospital, Milan, Italy
11. NeuroMuscular Omnicenter, Serena Onlus Foundation, Milan, Italy
12. Neuroalgology and Headache Unit, IRCCS Fondazione Istituto Neurologico "Carlo Besta", Milano, Italy
13. Department of Neurorehabilitation Sciences (P.T., M.C.), Casa Cura Policlinico, Milan, Italy
14. Department of Neurology and Institute of Experimental Neurology (INSPE), IRCCS San Raffaele Scientific Institute, Milan, Italy
15. Neurorehabilitation Unit/ALS Center, Scientific Clinical Institutes (ICS) Maugeri, IRCCS, Mistretta, Messina, Italy
16. Department of Neuroscience, S. Agostino-Estense Hospital, University of Modena and Reggio Emilia, Modena, Italy
17. Neurology Unit, School of Medicine and Surgery and NeuroMI, University of Milano-Bicocca, Monza, Italy
18. Department of Medical, Surgical, Neurological, Metabolic and Aging Sciences, University of Campania "Luigi Vanvitelli", Naples, Italy
19. Department of Medical, Surgical Neurological Metabolic and Aging Sciences, Second University of Naples, Naples, Italy
20. Department of Neurology, Azienda Ospedaliera San Francesco, Nuoro, Italy.
21. Division of Neurology, "A. Segni" Hospital, Ozieri, Italy
22. ALS Clinical Research Center, Bio. Ne. C., University of Palermo, Palermo, Italy.
23. Centro Clinico NEMO-Roma. Neurological Institute, Catholic University and I.C.O.M.M. Association for ALS Research, Rome, Italy
24. Department of Medicine, Azienda Complesso Ospedaliero, San Filippo Neri, Rome, Italy
25. Neurology Department, San Camillo Hospital, Rome, Italy
26. Department of Biomedical and Surgical Sciences, Section of Neurological, Psychiatric and Psychological Sciences, University of Ferrara, Ferrara, Italy
27. Department of Medical, Surgical and Neurological Sciences, University of Siena, Siena, Italy
28. Azienda Sanitaria Locale n. 2, Olbia-Tempio, Olbia, Italy
29. Rita Levi Montalcini Department of Neuroscience, Amyotrophic Lateral Sclerosis Center, University of Turin, Turin, Italy; Azienda Ospedaliero Universitaria Città della Salute e della Scienza, Turin, Italy
30. Molecular Genetics Unit, Department of Clinical Pathology, A.S.O. O.I.R.M.-S. Anna, 10126 Turin, Italy
31. Department of Neurology, Santa Chiara Hospital, Trento, Italy
32. Department of Neurological Rehabilitation, "Salvatore Maugeri" Clinical-Scientific Institute, Istituto di Ricovero e Cura a Carattere Scientifico, Veruno, Italy
33. Deceased

NYGC ALS Consortium


1. Center for Genomics of Neurodegenerative Diseases (CGND), New York Genome Center, New York, NY
2. Department of Neurology, University of Maryland School of Medicine, University of Maryland ALS Clinic, Baltimore, MD*
3. Cedars-Sinai Department of Biomedical Sciences, Board of Governors Regenerative Medicine Institute and Brain Program, Cedars-Sinai Medical Center, and Department of Medicine, University of California, Los Angeles, CA
4. Department of Biochemistry and Molecular Biology, Penn State Institute for Personalized Medicine, The Pennsylvania State University, Hershey, PA
5. Department of Neurology, The Pennsylvania State University, Hershey, PA
6. Department of Neurology, Henry Ford Health System, Detroit, MI
7. Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA
8. Department of Neurology, Center for Motor Neuron Biology and Disease, Institute for Genomic Medicine, Columbia University, New York, NY
9. Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA
10. Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD
11. Department of Neurogenetics, Academic Medical Centre, Amsterdam and Leiden University Medical Center, Leiden, The Netherlands
12. Department of Medicine, Lung Biology Center, University of California, San Francisco, CA
13. ALS Multidisciplinary Clinic, Neuromuscular Division, Department of Neurology, Harvard Medical School, and Neurological Clinical Research Institute, Massachusetts General Hospital, Boston, MA
14. Centre for Neuroscience and Trauma, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, and Department of Neurology, Basildon University Hospital, Basildon, United Kingdom
15. Institute of Neurology, National Hospital for Neurology and Neurosurgery, University College London, London, United Kingdom
16. The Jackson Laboratory, Bar Harbor, ME
17. Department of Psychiatry & Human Behavior, Department of Biological Chemistry, School of Medicine, and Department of Neurobiology and Behavior, School of Biological Sciences, University California, Irvine, CA
18. Taube/Koret Center for Neurodegenerative Disease Research, Roddenberry Center for Stem Cell Biology and Medicine, Gladstone Institute
19. Department of Neurology & Sensory Organs, University of Thessaly, Thessaly, Greece
20. Department of Neurology, Washington University in St. Louis, St. Louis, MO
21. Centre for Clinical Brain Sciences, Anne Rowling Regenerative Neurology Clinic, Euan MacDonald Centre for Motor Neurone Disease Research, University of Edinburgh, Edinburgh, United Kingdom
22. Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel
23. Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY
24. Center for Neurodegenerative Disorders, Department of Neurology, the Lewis Katz School of Medicine, Temple University, Philadelphia, PA
25. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
26. Computer Science and Systems Biology Program, Ann Romney Center for Neurological Diseases, Department of Neurology and Division of Genetics in Department of Medicine, Brigham and Women’s Hospital, Boston, MA
27. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA
28. Department of Anesthesiology, Stony Brook University, Stony Brook, NY
29. Section of Infections of the Nervous System, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD
Project MinE ALS Sequencing Consortium


1. Maurice Wohl Clinical Neuroscience Institute, King’s College London, Department of Basic and Clinical Neuroscience, London, UK.
2. Umeå University, Umeå, Sweden
3. Suna and Inan Kirac Foundation, Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey
4. Centre for MND Research, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW 2109, Australia
5. Universita degli Studi dei Torino, Turin, Italy
6. Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK
7. Université François-Rabelais, Tours
8. Université de Limoges, France
9. Instituto de Medicina Molecular, University of Lisbon
10. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
11. Tel-Aviv Medical Center, Tel-Aviv, Israel
12. Hospital Carlos III, Madrid, Spain
13. Hadassah University Hospital, Jerusalem, Israel
14. Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Republic of Ireland
15. Department of Neurology, Beaumont Hospital, Dublin, Republic of Ireland
16. Biostatistics Department, Harvard School of Public Health, Boston, Massachusetts, USA
17. Department of Biostatistics, IoPPN, King’s College London, London, US
18. Emory University, Atlanta, USA
19. Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA
20. Macquarie University, Sydney, Australia
21. SURFsara, Amsterdam, the Netherlands
22. Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland
23. University of Exeter Medical School, Exeter University, St Luke’s Campus, Magdalen Street, Exeter EX1 2LU, UK
24. Universidade de São Paulo, Brazil
25. KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium
26. VIB, Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium
27. University Hospitals Leuven, Department of Neurology, Leuven, Belgium
28. Hospital San Rafael, Madrid, Spain
29. Faculty of Medicine, University of Southampton, Southampton, UK
30. Biomedical Research Centre for Mental Health, IoPPN, King’s College London, London, UK
31. IRCCS Instituto Auxologico Italiano, Milan, Italy
32. Kantonspital St Gallen, St Gallen, Switzerland
33. Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW 2109, Australia

SLAGEN Consortium
Vincenzo Silani¹,², Nicola Ticozzi¹,², Cinzia Gellera³, Antonia Ratti¹,², Franco Taroni³, Giuseppe Lauria⁴, Federico Verde¹, Isabella Foghi⁵, Cinzia Tiloca¹, Giacomo P. Comi²,⁶, Gianni Sorarù⁷, Cristina Cereda⁸, Sandra D’Alfonso⁹, Lucia Corrado⁹, Fabiola De Marchi¹⁰, Stefania Corti²,⁶, Mauro Ceroni⁸,¹¹, Letizia Mazzini¹², Gabriele Siciliano¹², Massimiliano Filosto¹³, Maurizio Inghilleri¹⁴, Silvia Peverelli¹, Claudia Colombrita¹², Barbara Poletti¹, Luca Maderna¹, Roberto Del Bo²,⁶, Stella Gagliardi⁸, Giorgia Querin⁷, Cinzia Bertolin⁷, Viviana Pensato³, Barbara Castellotti³

1. Department of Neurology, IRCCS Istituto Auxologico Italiano, Milan, Italy
2. Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, Università degli Studi di Milano, Milan, Italy
3. Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan, Italy.
4. 3rd Neurology Unit, Motor Neuron Diseases Center, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milan, Italy.
5. Maurice Wohl Clinical Neuroscience Institute, King’s College London, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, London, UK.
6. Neurology Unit, IRCCS Foundation Ca’ Granda Ospedale Maggiore Policlinico, Milan, Italy.
7. Department of Neurosciences, University of Padova, Padova, Italy
8. Experimental Neurobiology Laboratory, ‘C. Mondino’ National Institute of Neurology Foundation, IRCCS, Pavia, Italy
9. Department of Health Sciences, University of Eastern Piedmont, Novara, Italy
10. Department of Neurology, Azienda Ospedaliera Universitaria Maggiore della Carità, Novara, Italy
11. Department of Neurological Sciences, University of Pavia, Pavia, Italy
12. Neurology Unit, Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy
13. Center for Neuromuscular Diseases and Neuropathies, Unit of Neurology ASST "Spedali Civili", University of Brescia, Brescia, Italy
14. Rare Neuromuscular Diseases Centre, Department of Neurology and Psychiatry, Sapienza University, Rome, Italy
Figure S1. Related to Figure 1; Workflow showing the quality control procedures applied to the present study. *increased to 10,031,630 when merged with the Van Rheenen et al dataset; Belg., Belgium; SNP, single nucleotide polymorphism; MAF, minor allele frequency, HWE, Hardy-Weinberg equilibrium; $R^2$, R-square value representing imputation precision; LNG, Laboratory of Neurogenetics.
Figure S2. Related to Figure 1; Multi-dimensional scaling plot of the 44,558 genotyped samples included in analysis compared to the HapMap populations.
Figure S3. Related to Figure 1; Quartile-Quartile plot of $P$-values from the meta-analysis based on logistic regression analysis. The black curve represents all SNPs, and the red curve represents SNPs after excluding variants within +/- 500 kilobases of the C9orf72 and the UNC13A loci. Raw genome inflation factor ($\lambda$) was 1.042 and adjusted $\lambda$ scaled to 1,000 cases and 1,000 controls was 1.001 based on the entire SNP dataset.
Figure S4. Related to Figure 3; Principal components analysis of samples included in the RVB analysis compared to the Human Diversity Panel. Ancestry filtering of the FALS discovery cohort was performed as follows: LASER was used to generate PCA coordinates for samples from the Human genome diversity panel (HGDP). Samples from the FALS discovery cohort were then mapped to this reference co-ordinate space. The discovery cohort was restricted to cases and controls occurring within 3 standard deviations of the mean for European HGDP samples along principal components 1-4.
Figure S5. Related to Figure 3; Quartile-Quartile plot of P values from the gene-based rare variant burden analysis of exome data. The genomic inflation factor ($\hat{\lambda} = 0.93$) was calculated based on the entire gene dataset.
**Figure S6. Related to Figure 3; Control-control analyses.** P values from RVB analysis of FALS cases versus controls (y-axis) are plotted against minimum P values from RVB analyses of candidate batch effects (x-axis). To assess the potential impact of batch effects, the sample cohort was divided into 28 pseudo case-control groups based on the sequencing center or associated dbGaP project. Loci showing possible association with non-ALS related batch effects are coloured light grey. No evidence of confounder bias was observed for KIF5A or previously reported ALS genes.
Figure S7. Related to Figure 3; Plot of variant call rates across the KIF5A protein-coding region in FALS versus controls analyzed by RVB testing.
Figure S8. Related to Figure 2; Principal components analysis of samples included in \textit{KIF5A} replication cohort.
Table S1. Related to Figure 1; Demographics and baseline characteristics of patients diagnosed with ALS and control individuals included in the GWAS analysis.

<table>
<thead>
<tr>
<th></th>
<th>US cases</th>
<th>US controls</th>
<th>Italian cases</th>
<th>Italian controls</th>
<th>UK cases</th>
<th>UK controls</th>
<th>French &amp; Belgian cases</th>
<th>French &amp; Belgian controls</th>
<th>Total cohort cases</th>
<th>Total cohort controls</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample number</strong></td>
<td>3,777</td>
<td>33,365</td>
<td>2,853</td>
<td>2,143</td>
<td>449</td>
<td>226</td>
<td>1,150</td>
<td>595</td>
<td>8,229</td>
<td>36,329</td>
</tr>
<tr>
<td><strong>Female (%)</strong></td>
<td>1,515</td>
<td>23,870</td>
<td>1,239</td>
<td>896</td>
<td>193</td>
<td>109</td>
<td>486</td>
<td>422</td>
<td>3,433</td>
<td>25,297</td>
</tr>
<tr>
<td>(40.1)</td>
<td>(43.4)</td>
<td>(41.8)</td>
<td>(43.0)</td>
<td>(48.2)</td>
<td>(42.3)</td>
<td>(70.9)</td>
<td>(41.7)</td>
<td>(69.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age (SD)</strong></td>
<td>58.1</td>
<td>64.2</td>
<td>61.8</td>
<td>50.6</td>
<td>60.3</td>
<td>57.0</td>
<td>60.5</td>
<td>66.9</td>
<td>59.8</td>
<td>63.4</td>
</tr>
<tr>
<td>(12.3)</td>
<td>(13.3)</td>
<td>(11.8)</td>
<td>(12.8)</td>
<td>(0.0)</td>
<td>(12.6)</td>
<td>(16.8)</td>
<td>(12.3)</td>
<td>(13.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>*<em>Bulbar-onset</em> (%)**</td>
<td>963</td>
<td>-</td>
<td>741</td>
<td>-</td>
<td>141</td>
<td>-</td>
<td>357</td>
<td>-</td>
<td>2,202</td>
<td>-</td>
</tr>
<tr>
<td>(25.5)</td>
<td>(26.0)</td>
<td>-</td>
<td>(31.4)</td>
<td>-</td>
<td>(31)</td>
<td>-</td>
<td>(26.8)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Family history† (%)</strong></td>
<td>458</td>
<td>-</td>
<td>248</td>
<td>-</td>
<td>54</td>
<td>-</td>
<td>195</td>
<td>-</td>
<td>955</td>
<td>-</td>
</tr>
<tr>
<td>(12.1)</td>
<td>(8.7)</td>
<td>-</td>
<td>(12.0)</td>
<td>-</td>
<td>(17.0)</td>
<td>-</td>
<td>(11.6)</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SD, standard deviation. *Data not available for site of symptom onset for 199 patients. †Data not available for familial history of 154 patients.
Table S2. Related to Figure 1; DbGaP studies contributing to the GWAS analysis.

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Study</th>
<th>Sample number</th>
<th>Females (%)</th>
<th>Average age (SD)</th>
<th>Genotyping platform</th>
<th>Ascertainment criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>phs000001</td>
<td>NEI Age-Related Eye Disease Study (AREDS)</td>
<td>1,644</td>
<td>959 (58.3)</td>
<td>68.2 (4.8)</td>
<td>HumanOmni2.5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000007</td>
<td>Framingham Cohort</td>
<td>1,298</td>
<td>718 (55.3)</td>
<td>75.7 (8.6)</td>
<td>HumanOmni5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000187</td>
<td>High Density SNP Association Analysis of Melanoma</td>
<td>1,027</td>
<td>414 (40.3)</td>
<td>51.3 (12.6)</td>
<td>HumanOmniExpress</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000196</td>
<td>CIDR: The NeuroGenetics Research Consortium Parkinson's Disease Study</td>
<td>10</td>
<td>6 (60)</td>
<td>74.3 (18.6)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000292</td>
<td>GENEVA Genetics of Early Onset Stroke (GEOS) Study</td>
<td>89</td>
<td>0 (0)</td>
<td>41.5 (6.4)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000304</td>
<td>Genes and Blood Clotting Study (GABC)</td>
<td>403</td>
<td>259 (64.3)</td>
<td>21.6 (3.3)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000315</td>
<td>Woman's Health Initiative (WHI GARNET)</td>
<td>4,206</td>
<td>4206 (100)</td>
<td>65.7 (6.9)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000368</td>
<td>Polycystic Ovary Syndrome Genetics (POLYGEN)</td>
<td>2,974</td>
<td>2973 (100)</td>
<td>46.8 (15.2)</td>
<td>HumanOmniExpress</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000372</td>
<td>Alzheimer's Disease Genetics Consortium Genome Wide Association Study</td>
<td>533</td>
<td>335 (62.9)</td>
<td>75.8 (9)</td>
<td>HumanOmniExpress</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000394</td>
<td>Autopsy-Confirmed Parkinson Disease GWAS Consortium (APDGC)</td>
<td>299</td>
<td>152 (50.8)</td>
<td>82.1 (12.6)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000397</td>
<td>NIA Long Life Family Study (LLFS)</td>
<td>1,804</td>
<td>957 (53)</td>
<td>65.9 (12.3)</td>
<td>HumanOmni2.5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000404</td>
<td>The Genetic Architecture of Smoking and Smoking Cessation</td>
<td>81</td>
<td>50 (61.7)</td>
<td>36.6 (5.9)</td>
<td>HumanOmni2.5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000421</td>
<td>A Genome-Wide Association Study of Fuchs' Endothelial Corneal Dystrophy</td>
<td>497</td>
<td>294 (59.2)</td>
<td>70.4 (10.2)</td>
<td>HumanOmni2.5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000428</td>
<td>Health and Retirement Study (HRS)</td>
<td>9,394</td>
<td>5437 (57.9)</td>
<td>68.4 (9.4)</td>
<td>HumanOmni2.5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000615</td>
<td>NINDS Stroke Genetics Network (SiGN)</td>
<td>743</td>
<td>416 (56)</td>
<td>56 (16.1)</td>
<td>HumanOmni5</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000675</td>
<td>GWAS on Selected WHI Hormone Trial European Americans</td>
<td>5,626</td>
<td>5626 (100)</td>
<td>68 (5.9)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000801</td>
<td>NCI Non-Hodgkin Lymphoma GWAS</td>
<td>1,544</td>
<td>790 (51.2)</td>
<td>58.4 (11.6)</td>
<td>HumanOmniExpress</td>
<td>Population controls</td>
</tr>
<tr>
<td>phs000869</td>
<td>Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study (BEAGESS)</td>
<td>1,174</td>
<td>271 (23.1)</td>
<td>61.3 (10.9)</td>
<td>HumanOmni1</td>
<td>Population controls</td>
</tr>
</tbody>
</table>
### Table S3. Related to Figure 1; SNPs achieving genome-wide significance in the GWAS analysis.

<table>
<thead>
<tr>
<th>SNP Information</th>
<th>Present Study (8,229 Cases / 36,329 Controls)</th>
<th>Van Rheenen et al. (12,577 Cases / 23,475 Controls)</th>
<th>Combined Discovery Set (20,806 Cases / 59,804 Controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novel Loci</strong></td>
<td><strong>Beta [SE]</strong></td>
<td><strong>OR [95% CI]</strong></td>
<td><strong>P</strong></td>
</tr>
<tr>
<td>rs117027576</td>
<td>0.373 [0.096]</td>
<td>1.45 [1.20-1.76]</td>
<td>1.1x10^{-4}</td>
</tr>
<tr>
<td>rs118082508</td>
<td>0.374 [0.096]</td>
<td>1.45 [1.20-1.76]</td>
<td>1.0x10^{-4}</td>
</tr>
<tr>
<td>rs113247976</td>
<td>0.381 [0.086]</td>
<td>1.46 [1.23-1.74]</td>
<td>9.2x10^{-6}</td>
</tr>
<tr>
<td>rs116900480</td>
<td>0.354 [0.083]</td>
<td>1.42 [1.21-1.68]</td>
<td>1.9x10^{-5}</td>
</tr>
<tr>
<td>rs142321490</td>
<td>0.357 [0.082]</td>
<td>1.43 [1.21-1.68]</td>
<td>1.5x10^{-5}</td>
</tr>
<tr>
<td><strong>Previously Published Loci</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10463311</td>
<td>-0.065 [0.024]</td>
<td>0.94 [0.89-0.98]</td>
<td>7.8x10^{-5}</td>
</tr>
<tr>
<td>rs3849943</td>
<td>-0.17 [0.024]</td>
<td>0.84 [0.80-0.88]</td>
<td>1.4x10^{-12}</td>
</tr>
<tr>
<td>rs74654358</td>
<td>0.182 [0.058]</td>
<td>1.20 [1.07-1.34]</td>
<td>1.6x10^{-3}</td>
</tr>
<tr>
<td>rs12973192</td>
<td>-0.149 [0.026]</td>
<td>0.86 [0.82-0.91]</td>
<td>1.3x10^{-8}</td>
</tr>
<tr>
<td>rs75087725</td>
<td>0.687 [0.162]</td>
<td>1.99 [1.44-2.75]</td>
<td>2.2x10^{-5}</td>
</tr>
</tbody>
</table>

Position is based on Human Genome Assembly build 37. Nearest gene or previously published gene names are included. Chr, chromosome; SE, standard error; OR, odds ratio; 95% CI, 95% confidence interval; *, rs113247976 represents the p.Pro986Leu variant in KIF5A (NM_004984.2).
Table S4. Related to Figure 1; Suggestive SNPs with $P$ values less than 5.0x10-7 in the GWAS analyses.

<table>
<thead>
<tr>
<th>SNP Information</th>
<th>Present Study (8,229 Cases / 36,329 Controls)</th>
<th>Van Rheenen et al. (12,577 Cases / 23,475 Controls)</th>
<th>Combined Discovery Set (20,806 Cases / 59,804 Controls)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP</td>
<td>Case MAF</td>
<td>Control MAF</td>
<td>Case MAF</td>
</tr>
<tr>
<td>SNP</td>
<td>Chr</td>
<td>Position</td>
<td>OR [95% CI]</td>
</tr>
<tr>
<td>rs17070492</td>
<td>8</td>
<td>2,420,855</td>
<td>LOC101927815</td>
</tr>
<tr>
<td>rs10139154</td>
<td>14</td>
<td>31,147,498</td>
<td>SCFD1</td>
</tr>
<tr>
<td>rs10143310</td>
<td>14</td>
<td>92,540,381</td>
<td>ATXN3</td>
</tr>
<tr>
<td>rs9901522</td>
<td>17</td>
<td>14,673,934</td>
<td>PMP22</td>
</tr>
</tbody>
</table>
Table S5. Related to Figure 3; DbGaP/EGA studies contributing to the RVB analysis.

<table>
<thead>
<tr>
<th>Accession Number</th>
<th>Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>phs000179</td>
<td>Genetic Epidemiology of COPD (COPDGene)</td>
</tr>
<tr>
<td>phs000254</td>
<td>NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Cystic Fibrosis)</td>
</tr>
<tr>
<td>phs000281</td>
<td>NHLBI GO-ESP: Women's Health Initiative Exome Sequencing Project (WHI) - WHISP</td>
</tr>
<tr>
<td>phs000290</td>
<td>NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Pulmonary Arterial Hypertension)</td>
</tr>
<tr>
<td>phs000291</td>
<td>NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Lung Health Study of COPD)</td>
</tr>
<tr>
<td>phs000296</td>
<td>NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (COPDGene)</td>
</tr>
<tr>
<td>phs000307</td>
<td>NHLBI Framingham Heart Study Allelic Spectrum Project</td>
</tr>
<tr>
<td>phs000347</td>
<td>NHLBI GO-ESP: Family Studies (Aortic Disease)</td>
</tr>
<tr>
<td>phs000354</td>
<td>NHLBI GO-ESP Family Studies: Pulmonary Arterial Hypertension</td>
</tr>
<tr>
<td>phs000362</td>
<td>NHLBI GO-ESP: Family Studies: (Familial Atrial Fibrillation)</td>
</tr>
<tr>
<td>phs000398</td>
<td>NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (ARIC)</td>
</tr>
<tr>
<td>phs000400</td>
<td>NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (CHS)</td>
</tr>
<tr>
<td>phs000401</td>
<td>NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (FHS)</td>
</tr>
<tr>
<td>phs000402</td>
<td>NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (JHS)</td>
</tr>
<tr>
<td>phs000403</td>
<td>NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (MESA)</td>
</tr>
<tr>
<td>phs000422</td>
<td>NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Asthma)</td>
</tr>
<tr>
<td>phs000498</td>
<td>Jackson Heart Study Allelic Spectrum Project</td>
</tr>
<tr>
<td>phs000518</td>
<td>NHLBI GO-ESP Family Studies: Idiopathic Bronchiectasis</td>
</tr>
<tr>
<td>phs000572</td>
<td>Alzheimer's Disease Sequencing Project (ADSP)</td>
</tr>
<tr>
<td>phs000652</td>
<td>NHLBI GO-ESP: Family Studies (Hematological Cancers)</td>
</tr>
<tr>
<td>phs000651</td>
<td>Building on GWAS: the U.S. CHARGE consortium - Sequencing (CHARGE-S): FHS</td>
</tr>
<tr>
<td>phs000667</td>
<td>Building on GWAS for NHLBI-Diseases: The U.S. CHARGE Consortium - Sequencing (CHARGE-S): CHS</td>
</tr>
<tr>
<td>phs000668</td>
<td>Building on GWAS: the U.S. CHARGE consortium - Sequencing (CHARGE-S): ARIC</td>
</tr>
<tr>
<td>phs000744</td>
<td>Yale Center for Mendelian Genomics (Y CMG)</td>
</tr>
<tr>
<td>phs000806</td>
<td>MIGen_ExS: Ottawa Heart Study</td>
</tr>
<tr>
<td>phs000814</td>
<td>MIGen_ExS: Italian Atherosclerosis Thrombosis and Vascular Biology</td>
</tr>
<tr>
<td>phs000908</td>
<td>Identification of Rare Variants in PD through Whole Exome Sequencing</td>
</tr>
<tr>
<td>phs000917</td>
<td>MIGen_ExS: PROMIS</td>
</tr>
<tr>
<td>phs001000</td>
<td>MIGen_ExS: U. of Leicester</td>
</tr>
<tr>
<td>phs001101</td>
<td>MIGen_ExS: MDC</td>
</tr>
<tr>
<td>EGAO00000000079</td>
<td>UK10K</td>
</tr>
<tr>
<td>phs000101</td>
<td>NIH Exome Sequencing of Familial Amyotrophic Lateral Sclerosis Project</td>
</tr>
</tbody>
</table>

Sample number | Females (%) |
--- | --- |
2 | 100% |
238 | 49.6% |
1904 | 100% |
73 | 82.2% |
332 | 37% |
285 | 52.6% |
1317 | 51.6% |
29 | 34.5% |
9 | 88.9% |
12 | 16.7% |
800 | 54.6% |
186 | 28% |
348 | 36.8% |
296 | 58.8% |
259 | 45.2% |
189 | 65.1% |
1629 | 63.8% |
24 | 70.8% |
4655 | 58.8% |
19 | 36.8% |
550 | 61.5% |
1209 | 52.9% |
5497 | 58.5% |
1944 | 44.7% |
1966 | 33.1% |
3591 | 11.3% |
105 | 66.7% |
7298 | 17.9% |
1081 | 0% |
1075 | 44.7% |
4062 | 65% |
201 | 45% |
Table S6. Related to Figure 2, 3; Quality control filtering of the FALS discovery and *KIF5A* replication cohorts.

**FALS discovery cohort**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Sample Set</td>
<td>1,463</td>
<td>41,410</td>
</tr>
<tr>
<td>Post HGDP Continental Ancestry Filter</td>
<td>1,397</td>
<td>24,563</td>
</tr>
<tr>
<td>Post Call Rate Filter</td>
<td>1,331</td>
<td>20,789</td>
</tr>
<tr>
<td>Post Heterozygosity Filter</td>
<td>1,319</td>
<td>20,664</td>
</tr>
<tr>
<td>Post Relatedness Filter</td>
<td>1,138</td>
<td>19,494</td>
</tr>
</tbody>
</table>

**rs113247976 replication cohort (FALS discovery + ALS WXS/WGS replication cohort)**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Sample Set</td>
<td>12,180*</td>
<td>21,533**</td>
</tr>
<tr>
<td>Post Call Rate Filter</td>
<td>11,916</td>
<td>21,050</td>
</tr>
<tr>
<td>Post Heterozygosity Filter</td>
<td>11,721</td>
<td>21,028</td>
</tr>
<tr>
<td>Post Ancestry Filter (PCA)</td>
<td>11,373</td>
<td>21,009</td>
</tr>
<tr>
<td>Post Relatedness &amp; GWAS Checksum Filter</td>
<td>4,160</td>
<td>18,650</td>
</tr>
</tbody>
</table>
* All 1,138 FALS passing QC in FALS discovery cohort + 11,042 additional ALS WXS/WGS cases
** All 19,494 controls passing QC in FALS discovery cohort + 2,039 additional WXS/WGS controls

**LOF screen (ALS WXS/WGS replication cohort)**

<table>
<thead>
<tr>
<th>Cohort</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Sample Set</td>
<td>11,042*</td>
<td>2,039**</td>
</tr>
<tr>
<td>Post Call Rate Filter</td>
<td>10,741</td>
<td>2,039</td>
</tr>
<tr>
<td>Post Heterozygosity Filter</td>
<td>10,549</td>
<td>2,026</td>
</tr>
<tr>
<td>Post Ancestry Filter (PCA)</td>
<td>10,201</td>
<td>2,008</td>
</tr>
<tr>
<td>Post Relatedness</td>
<td>9,046</td>
<td>1,955</td>
</tr>
</tbody>
</table>
* 11,042 additional ALS WXS/WGS cases not included in FALS discovery cohort
** 2,039 additional WXS/WGS controls not included in FALS discovery cohort

See Experimental Procedures for further details on filtering parameters.
Table S7. Related to Figure 3; RVB analysis according to mutation type across KIF5A and within gene sub-domains.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>FALS</th>
<th>Control</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense - Full CDS</td>
<td>9 (0.79%)</td>
<td>80 (0.41%)</td>
<td>1.93 (0.915-3.60)</td>
<td>8.09x10^{-2}</td>
</tr>
<tr>
<td>Missense - Motor Domain</td>
<td>3 (0.26%)</td>
<td>18 (0.09%)</td>
<td>3.27 (0.86-9.25)</td>
<td>7.74x10^{-2}</td>
</tr>
<tr>
<td>Missense - Microtubule Binding Domain</td>
<td>2 (0.18%)</td>
<td>8 (0.04%)</td>
<td>5.07 (0.95-18.52)</td>
<td>5.57x10^{-2}</td>
</tr>
<tr>
<td>Missense - Coiled-Coil Domain</td>
<td>3 (0.26%)</td>
<td>55 (0.28%)</td>
<td>1.01 (0.28-2.60)</td>
<td>9.83x10^{-1}</td>
</tr>
<tr>
<td>Missense - C-Terminal Domain</td>
<td>3 (0.26%)</td>
<td>7 (0.04%)</td>
<td>7.23 (1.74-24.55)</td>
<td>9.41x10^{-3}</td>
</tr>
<tr>
<td>Loss of Function</td>
<td>6 (0.53%)</td>
<td>3 (0.02%)</td>
<td>32.07 (9.05-135.27)</td>
<td>5.55x10^{-7}</td>
</tr>
<tr>
<td>Loss of Function (including frameshifts)</td>
<td>8 (0.70%)</td>
<td>3 (0.02%)</td>
<td>41.16 (12.61-167.57)</td>
<td>3.77x10^{-9}</td>
</tr>
</tbody>
</table>

FALS, familial ALS; OR, odds ratio; 95% CI, 95% confidence interval; CDS, coding sequence
Table S8. Related to Figure 3; Clinical information of probands and relatives carrying KIF5A LOF variants.

<table>
<thead>
<tr>
<th>Position</th>
<th>Variant</th>
<th>Relation to Proband</th>
<th>DNA Available</th>
<th>Exon</th>
<th>cDNA</th>
<th>Description</th>
<th>Gender</th>
<th>Age of Onset (years)</th>
<th>Site of Onset</th>
<th>Survival (months)</th>
<th>Alive</th>
</tr>
</thead>
<tbody>
<tr>
<td>57,975,729</td>
<td>GA&gt;A</td>
<td>Proband</td>
<td>Y</td>
<td>26</td>
<td>c.2987delA</td>
<td>p.Asp996fs</td>
<td>M</td>
<td>45</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>57,976,382</td>
<td>C&gt;T</td>
<td>Proband</td>
<td>Y</td>
<td>27</td>
<td>c.2993-3C&gt;T</td>
<td>5' Splice Junction</td>
<td>M</td>
<td>29</td>
<td>L</td>
<td>&gt;264</td>
<td>Y</td>
</tr>
<tr>
<td>57,976,382</td>
<td>C&gt;T</td>
<td>Sister</td>
<td>Y</td>
<td>27</td>
<td>c.2993-3C&gt;T</td>
<td>5' Splice Junction</td>
<td>F</td>
<td>52</td>
<td>L</td>
<td>84</td>
<td>N</td>
</tr>
<tr>
<td>57,976,382</td>
<td>C&gt;T</td>
<td>Brother</td>
<td>Y</td>
<td>27</td>
<td>c.2993-3C&gt;T</td>
<td>5' Splice Junction</td>
<td>M</td>
<td>18</td>
<td>L</td>
<td>324</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brother</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>n/a</td>
<td>L</td>
<td>n/a</td>
<td>N</td>
</tr>
<tr>
<td>57,975,731</td>
<td>CA&gt;C</td>
<td>Sporadic</td>
<td>Y</td>
<td>26</td>
<td>c.2989delA</td>
<td>p.Asn997fs</td>
<td>F</td>
<td>50</td>
<td>L</td>
<td>&gt;96</td>
<td>Y</td>
</tr>
<tr>
<td>57,976,384</td>
<td>G&gt;A</td>
<td>Sporadic</td>
<td>N</td>
<td>27</td>
<td>c.2993-1G&gt;A</td>
<td>5' Splice Junction</td>
<td>n/a</td>
<td>52</td>
<td>B</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>57,976,385</td>
<td>GA&gt;G</td>
<td>Proband</td>
<td>Y</td>
<td>27</td>
<td>c.2996delA</td>
<td>p.Asn999fs</td>
<td>M</td>
<td>42</td>
<td>L</td>
<td>&gt;12</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Brother</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>M</td>
<td>38</td>
<td>n/a</td>
<td>24</td>
<td>N</td>
</tr>
<tr>
<td>57,976,412</td>
<td>G&gt;A</td>
<td>Proband</td>
<td>Y</td>
<td>27</td>
<td>c.3020G&gt;A</td>
<td>p.Arg1007Lys</td>
<td>F</td>
<td>50</td>
<td>n/a</td>
<td>&gt;240</td>
<td>Y</td>
</tr>
<tr>
<td>57,976,413</td>
<td>G&gt;A</td>
<td>Proband</td>
<td>Y</td>
<td>27</td>
<td>c.3020+1G&gt;A</td>
<td>3' Splice Junction</td>
<td>M</td>
<td>45</td>
<td>B</td>
<td>&gt;220</td>
<td>Y</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Parent</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td>47</td>
<td>n/a</td>
<td>156</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncle/Aunt</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td>57</td>
<td>n/a</td>
<td>144</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncle/Aunt</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td>55</td>
<td>n/a</td>
<td>121</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Uncle/Aunt</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>n/a</td>
<td>46</td>
<td>n/a</td>
<td>24</td>
<td>N</td>
</tr>
<tr>
<td>57,976,414</td>
<td>T&gt;A</td>
<td>Proband</td>
<td>Y</td>
<td>27</td>
<td>c.3020+2T&gt;A</td>
<td>3' Splice Junction</td>
<td>M</td>
<td>46</td>
<td>B</td>
<td>124</td>
<td>N</td>
</tr>
<tr>
<td>57,976,414</td>
<td>T&gt;A</td>
<td>Brother</td>
<td>Y</td>
<td>27</td>
<td>c.3020+2T&gt;A</td>
<td>3' Splice Junction</td>
<td>M</td>
<td>48</td>
<td>L</td>
<td>117</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mother</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td>F</td>
<td>35</td>
<td>L</td>
<td>144</td>
<td>N</td>
</tr>
<tr>
<td>57,976,415</td>
<td>A&gt;G</td>
<td>Proband</td>
<td>Y</td>
<td>27</td>
<td>c.3020+3A&gt;G</td>
<td>3' Splice Junction</td>
<td>M</td>
<td>50</td>
<td>B</td>
<td>54</td>
<td>N</td>
</tr>
</tbody>
</table>

All mutations were heterozygous; Genomic coordinates are based on Human Genome Assembly build 37; Protein change is based on transcript NM_004984.3; n/a, not applicable or not available;
## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antibodies</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bacterial and Virus Strains</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biological Samples</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chemicals, Peptides, and Recombinant Proteins</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 15596026</td>
</tr>
<tr>
<td>TRIzol Reagent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA-to-cDNA Kit</td>
<td>Applied Biosystems</td>
<td>Cat# 4368814</td>
</tr>
<tr>
<td>RNase Inhibitor</td>
<td>Applied Biosystems</td>
<td>Cat# 4374966</td>
</tr>
<tr>
<td>OneTaq Hot Start DNA Polymerase</td>
<td>New England BioLabs</td>
<td>Cat# M0481S</td>
</tr>
<tr>
<td>Critical Commercial Assays</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product Name</td>
<td>Manufacturer</td>
<td>Cat#</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>HumanOmniExpress-24 DNA Analysis Kit</td>
<td>Illumina</td>
<td>Cat# WG-312-3007</td>
</tr>
<tr>
<td>TruSeq Exome Enrichment Kit</td>
<td>Illumina</td>
<td>Cat# FC-121-1096</td>
</tr>
<tr>
<td>TruSeq PE Cluster Kit</td>
<td>Illumina</td>
<td>Cat# PE-401-3001</td>
</tr>
<tr>
<td>TruSeq SBS Kit</td>
<td>Illumina</td>
<td>Cat# FC-401-3001</td>
</tr>
<tr>
<td>TrueSeq DNA PCR-free Kit</td>
<td>Illumina</td>
<td>Cat# 20000902</td>
</tr>
<tr>
<td>HiSeq X Ten Reagent Kit</td>
<td>Illumina</td>
<td>Cat# FC-501-2501</td>
</tr>
</tbody>
</table>

**Deposited Data**

**Experimental Models: Cell Lines**

**Experimental Models: Organisms/Strains**

**Oligonucleotides**
<table>
<thead>
<tr>
<th></th>
<th>F1 primer sequence: CAGTGGAGCCACATCTTCTG</th>
<th>Operon Technologies</th>
<th>NA</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1 primer sequence: TCTCTTGGTGGAGGGGAAA</td>
<td>Operon Technologies</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td>F2 primer sequence: CCAACATGGACAATGGAGTGA</td>
<td>Operon Technologies</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

Recombinant DNA


Software and Algorithms

<table>
<thead>
<tr>
<th>Software</th>
<th>Reference</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASSEDA</td>
<td>Tompson et al., 2007</td>
<td><a href="http://www.cytognix.com/?post_type=duka&amp;p=2670">http://www.cytognix.com/?post_type=duka&amp;p=2670</a></td>
</tr>
<tr>
<td>BWA</td>
<td>Wellcome Trust</td>
<td><a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a></td>
</tr>
<tr>
<td></td>
<td>Sanger Institute</td>
<td></td>
</tr>
<tr>
<td>GenABEL</td>
<td>The GenABEL Project</td>
<td><a href="http://www.genabel.org">http://www.genabel.org</a></td>
</tr>
<tr>
<td>GATK</td>
<td>Broad Institute</td>
<td><a href="https://software.broadinstitute.org/gatk/">https://software.broadinstitute.org/gatk/</a></td>
</tr>
<tr>
<td>id_geno_checksum</td>
<td>Broad Institute</td>
<td><a href="https://personal.broadinstitute.org/sripke/share_links/checksums_download/">https://personal.broadinstitute.org/sripke/share_links/checksums_download/</a></td>
</tr>
<tr>
<td>KING</td>
<td>Manichaikul et al., 2010</td>
<td><a href="http://people.virginia.edu/~wc9c/KING/">http://people.virginia.edu/~wc9c/KING/</a></td>
</tr>
<tr>
<td>LASER</td>
<td>University of Michigan</td>
<td><a href="http://csg.sph.umich.edu/chaolong/LASER/">http://csg.sph.umich.edu/chaolong/LASER/</a></td>
</tr>
<tr>
<td>TOOL</td>
<td>DESCRIPTION</td>
<td>PUBLICATION YEAR</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Mach2dat</td>
<td>Marchini and Howie, 2010</td>
<td><a href="https://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output">https://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output</a></td>
</tr>
<tr>
<td>METAL</td>
<td>Willer et al., 2010</td>
<td><a href="http://csg.sph.umich.edu/abecasis/metal/index.html">http://csg.sph.umich.edu/abecasis/metal/index.html</a></td>
</tr>
<tr>
<td>Minimac3</td>
<td>Das et al., 2016</td>
<td><a href="https://genome.sph.umich.edu/wiki/Minimac3">https://genome.sph.umich.edu/wiki/Minimac3</a></td>
</tr>
<tr>
<td>PLINK</td>
<td>Chang et al., 2015</td>
<td><a href="http://zzz.bwh.harvard.edu/plink/">http://zzz.bwh.harvard.edu/plink/</a></td>
</tr>
<tr>
<td>R</td>
<td>R Core Team</td>
<td><a href="https://www.r-project.org">https://www.r-project.org</a></td>
</tr>
<tr>
<td>SHAPEIT</td>
<td>Delaneau et al., 2013</td>
<td><a href="https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html">https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html</a></td>
</tr>
<tr>
<td>dbNSFP</td>
<td>Liu et al., 2013</td>
<td><a href="http://varianttools.sourceforge.net/Annotation/DbNSFP">http://varianttools.sourceforge.net/Annotation/DbNSFP</a></td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Highlights**

- Loss of function mutations in *KIF5A* are a cause of amyotrophic lateral sclerosis
- ALS-associated *KIF5A* mutations are distinct from HSP and CMT mutations in *KIF5A*
- Identification of *KIF5A* highlights role of cytoskeleton in ALS pathogenesis

**eTOC**

Using large-scale genome-wide association study and exome sequencing, we identified *KIF5A* as a novel gene associated with ALS. Our data broaden the phenotype resulting from mutations in *KIF5A* and highlight the importance of cytoskeletal defects in the pathogenesis of ALS.