Genome-wide Analyses Identify KIF5A as a Novel ALS Gene

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 the role of cytoskeleton in ALS pathogenesis

Authors

Aude Nicolas, Kevin P. Kenna, Alan E. Renton, ..., Christopher E. Shaw, Bryan J. Traynor, John E. Landers

Correspondence

bryan.traynor@nih.gov (B.J.T.), john.landers@umassmed.edu (J.E.L.)

In Brief

Using a 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.
Genome-wide Analyses Identify KIF5A as a Novel ALS Gene

Aude Nicolas,1,137 Kevin P. Kenna,2,137 Alan E. Renton,1,3,4,137 Nicola Ticozzi,5,6,137 Faraz Faghri,7,8,137 Ruth Chia,1,137 Janice A. Dominov,2 Brendan J. Kenna,2 Mike A. Nalls,7,8 Pamela Keagle,2 Alberto M. Rivera,1 Wouter van Rheenen,10 Natalie A. Murphy,1 Joke J.F.A. van Vugt,10 Joshua T. Geiger,11 Rick A. Van der Spek,10 Hannah A. Pliner,1 Shankaracharya,2 Bradley N. Smith,12 Giuseppe Marangi,1,13 Simon D. Topp,12 Yevgeniya Abramzon,1,4 Athina Soragia Gkazi,12 John D. Eicher,15 Aoife Kenna,2 ITALSGEN Consortium, Gabriele Mora,16 Andrea Calvo,17 Letizia Mazzini,16 Nilo Riva,19 Jessica Mandrioli,20 Claudia Caponnetto,21 Stefania Battistini,22 Paolo Volanti,16 Vincenzo La Bella,23 Francesca L. Conforti,24 Giuseppe Borghero,25 Sonia Messina,26 Isabella L. Simone,27 Francesca Trojci,28 Fabrizio Salvi,29 Francesco O. Logullo,30 Sandra D’Alfonso,31 Lucia Corrado,31 Margherita Capasso,32 Luigi Ferrucci,33 Genomic Translation for ALS Care (GTAC) Consortium, Cristiane de Araujo Martins Moreno,34 Sitharthan Kamalakaran,35 David B. Goldstein,35 ALS Sequencing Consortium, Aaron D. Gitler,36 Tim Harris,37 Richard M. Myers,38 NYGC ALS Consortium, Hemali Phatnani,39 Rajeeva Lochan Musunuri,40 Uday Shankar Evani,40 Avinash Abhyankar,40 Michael C. Zody,40 Answer ALS Foundation, Julia Kaye,41 Steven Finkbeiner,41,42

1Neuromuscular Diseases Research Section, Laboratory of Neurogenetics, National Institute on Aging, NIH, Porter Neuroscience Research Center, Bethesda, MD 20892, USA
2Department of Neurology, University of Massachusetts Medical School, Worcester, MA 01605, USA
3Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
4Ronald M. Loeb Center for Alzheimer’s Disease, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
5Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan, Italy
6Department of Pathophysiology and Transplantation, “Dino Ferrari” Center – Università degli Studi di Milano, Milan 20122, Italy
7Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, NIH, Porter Neuroscience Research Center, Bethesda, MD 20892, USA
8Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA
9Data Tecnica International, Glen Echo, MD, USA
10Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands
11Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD 20892, USA
12Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, King’s College London, London SE5 9RS, UK
13Institute of Genomic Medicine, Catholic University, Roma, Italy
14Sobell Department of Motor Neuroscience and Movement Disorders, University College London, Institute of Neurology, London, UK
15Genetics and Pharmacogenomics, MRL, Merck & Co., Inc., Boston, MA 02115, USA
16ALS Center, Salvatore Maugeri Foundation, IRCCS, Mistretta, Messina, Italy
17“Rita Levi Montalcini” Department of Neuroscience, University of Turin, Turin, Italy
18“Maggiore della Carità” University Hospital, Novara, Italy

(Affiliations continued on next page)

SUMMARY

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: 05400) is a neurodegenerative disorder clinically characterized by rapidly progressive muscle weakness and death due to respiratory failure, typically within 2 to 4 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...

(Author list continued on next page)

(Affiliations continued on next page)

Enfermedades Raras (CIBERER U-723), Madrid, Spain
*Correspondence: bryan.traynor@nih.gov (B.J.T.), john.landers@umassmed.edu (J.E.L.)

100KU Leuven – University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium
101VIB, Center for Brain and Disease Research, Laboratory of Neurobiology, Leuven, Belgium
102Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA
103Institute of Human Genetics, Technische Universität München, Munich, Germany
104Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
105Munich Cluster for Systems Neurology (SyNergy), Munich, Germany
106Institute of Physiology, Institute of Molecular Medicine, University of Lisbon, Lisbon, Portugal
107Department of Neurosciences, Hospital de Santa Maria-CHLN, Lisbon, Portugal
108SURF-sara, Amsterdam, the Netherlands
109Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland
110Neuromuscular Diseases Center/ALS Clinic, Kantonsspital St. Gallen, St. Gallen, Switzerland
111Department of Neurology, Mayo Clinic Florida, Jacksonville, FL 32224, USA
112Department of Neuroscience, Mayo Clinic, Jacksonville, FL, USA
113Faculty of Medicine, University of Southampton, Southampton, UK
114Suna and Inan Kırac Foundation, Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey
115ALS Unit/Neurology, Hospital San Rafael, Madrid, Spain
116Department of Neurology, Tel-Aviv Sourasky Medical Centre, Tel-Aviv, Israel
117Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
118Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Republic of Ireland
119Centre for MND Research, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW 2109, Australia
120ANZAC Research Institute, Concord Hospital, University of Sydney, Sydney, NSW 2139, Australia
121Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada
122Unidad 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
123Tanz Centre for Research of Neurodegenerative Diseases, Division of Neurology, Department of Medicine, University of Toronto, Toronto, ON M5S 3H2, Canada
124Division of Neurology, Department of Internal Medicine, Sunnybrook Health Sciences Centre, University of Toronto, Toronto, ON M4N 3M5, Canada
125Department of Neurology, Penn State Hershey Medical Center, Hershey, PA, USA
126Department of Neurology, University of Michigan, Ann Arbor, MI, USA
127Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico “Carlo Besta,” Milan 20133, Italy
128University Hospitals Leuven, Department of Neurology, Leuven, Belgium
129Centro Clínico NeMo, Institute of Neurology, Umeå University, Umeå SE-90185, Sweden
130NEuroMuscular Omnicenter (NEMO), Serena Onlus Foundation, Milan, Italy
131ALS Center, CHU Gui de Chauliac, University of Montpellier, Montpellier, France
132Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
133Institute of Human Genetics, Technische Universität München, Munich, Germany
134Institute of Human Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany
135Genetics and Pharmacogenomics, MRL, Merck & Co., Inc., West Point, PA 19486, USA
136Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA, USA
137NeuroMuscular Omnicenter (NEMO), Serena Onlus Foundation, Milan, Italy
138Lead Contact
*Correspondence: bryan.traynor@nih.gov (B.J.T.), john.landers@umassmed.edu (J.E.L.)

https://doi.org/10.1016/j.neuron.2018.02.027

(Approximately 10% of ALS cases 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 FALS cases and about 10% of SALS 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

(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).
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 frontotemporal dementia (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 toward 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 C9orf72, TDP-43, FUS, HNRNPA1, and MATR3), (2) protein homeostasis (UBQLN2, VCP, OPTN, and VAPB), and (3) cytoskeletal dynamics (PFN1, TUBA4A, and 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 TBK1 is known to interact with the product of ALS-associated gene 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 NEK1, identified through a rare variant burden analysis (RVB) of index FALS (i.e., one affected sample per family), is a known binding partner of C2orf72, an ALS risk factor found through genome-wide association studies (GWASs) (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 that is becoming increasingly implicated as a contributing factor in ALS and other neurodegenerative diseases (Coppede and Migliore, 2015; Lopez-Gonzalez et al., 2016; Madabhushi et al., 2014; Wang et al., 2013).

RESULTS

GWASs 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 (Tables S1 and 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 STAR Methods and 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 ($\lambda_{HC} = 1.001$;
Figure S3). SNPs achieving genome-wide significance (p < 5.0 × 10^{-8}) are listed in Tables 1 and S3 and suggestive loci with SNPs associated at p < 5.0 × 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, and 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 spanning a region of several hundred kilobases (Table 1; Figure 1B). 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 × 10^{-10}, odds ratio [OR] = 1.38, 95% CI = 1.24–1.53; Figure 2). 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 reference allele; Figure 2).

Calculations based on our cohort size as well as the OR and allele frequency of rs113247976 result in an ~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 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 and 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).

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.2993-3G>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 validated all identified LOF variants containing samples in the discovery set and affected relatives.

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 affect splicing of exon 27, which 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 on the 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; Figures 4B and 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 and 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 × 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**

For 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 (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.
Table 1. SNPs Achieving Genome-wide Significance in the Discovery GWAS

<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>Chr</td>
<td>Position</td>
<td>Gene</td>
</tr>
<tr>
<td>Novel Loci</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs117027576</td>
<td>12</td>
<td>57,316,603</td>
<td>KIF5A</td>
</tr>
<tr>
<td>rs118082508</td>
<td>12</td>
<td>57,318,819</td>
<td>KIF5A</td>
</tr>
<tr>
<td>rs113247976*</td>
<td>12</td>
<td>57,975,700</td>
<td>KIF5A</td>
</tr>
<tr>
<td>rs116900480</td>
<td>12</td>
<td>58,656,105</td>
<td>KIF5A</td>
</tr>
<tr>
<td>rs142321490</td>
<td>12</td>
<td>58,676,132</td>
<td>KIF5A</td>
</tr>
<tr>
<td>Previously Published Loci</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10463311</td>
<td>5</td>
<td>150,410,835</td>
<td>TNIP1</td>
</tr>
<tr>
<td>rs3849943</td>
<td>9</td>
<td>27,543,382</td>
<td>C9orf72</td>
</tr>
<tr>
<td>rs74654358</td>
<td>12</td>
<td>64,881,967</td>
<td>TBK1</td>
</tr>
<tr>
<td>rs12973192</td>
<td>19</td>
<td>17,753,239</td>
<td>UNCI3A</td>
</tr>
<tr>
<td>rs75087725</td>
<td>21</td>
<td>45,753,117</td>
<td>C2orf2</td>
</tr>
</tbody>
</table>

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 (GenBank: 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>OR (95% CI)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>KIF5A</td>
<td>3.07 (2.02 - 5.07)</td>
<td>2.02 x 10⁻⁴</td>
</tr>
<tr>
<td>TBK1</td>
<td>1.51 (1.31 - 1.72)</td>
<td>2.03 x 10⁻⁵</td>
</tr>
<tr>
<td>NEK1</td>
<td>1.21 (1.09 - 1.33)</td>
<td>8.07 x 10⁻⁴</td>
</tr>
<tr>
<td>CALHM2</td>
<td>1.23 (1.13 - 1.32)</td>
<td>1.38 x 10⁻⁴</td>
</tr>
<tr>
<td>COL14A1</td>
<td>1.55 (1.37 - 1.74)</td>
<td>1.79 x 10⁻⁵</td>
</tr>
<tr>
<td>AK1</td>
<td>1.22 (1.10 - 1.35)</td>
<td>4.84 x 10⁻⁴</td>
</tr>
<tr>
<td>ATN</td>
<td>1.00 (0.90 - 1.10)</td>
<td>6.03 x 10⁻⁴</td>
</tr>
<tr>
<td>VLDLR</td>
<td>1.20 (1.09 - 1.32)</td>
<td>1.79 x 10⁻⁴</td>
</tr>
<tr>
<td>FUS</td>
<td>1.15 (0.98 - 1.34)</td>
<td>2.54 x 10⁻⁴</td>
</tr>
<tr>
<td>ZMYND12</td>
<td>1.20 (1.09 - 1.32)</td>
<td>1.79 x 10⁻⁴</td>
</tr>
</tbody>
</table>

Figure 2. Discovery and Replication of the Association of the KIF5A p.Pro886Leu (rs113247976) Variant with ALS

Analysis of the p.Pro886Leu (rs113247976) variant within each of the described cohorts is shown. Allelic association for all subcohorts was 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 gray line denoting the OR estimated under the meta-analysis, and with the width of the horizontal lines and the diamonds corresponding to 95% confidence intervals.

STAR Methods; Figure S8). This included non-overlapping samples from our RVB (673 FALS and 17,696 controls). Analysis of the cohort revealed an allele frequency of 1.78% in cases and 1.78% in controls (rs113247976, p = 3.82 x 10⁻⁴), thereby replicating the association of the original GWAS. A meta-analysis of the GWAS and replication cohort (n = 24,965 cases and 78,454 controls) yielded a highly significant p value of 7.09 x 10⁻¹³ (OR = 1.39, 95% CI = 1.33–1.45; Figure 2). These results support the association of KIF5A p.Pro886Leu with ALS. However, at this point we cannot definitively 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. 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 FALS patients from our RVB; however, a comparison of 240,715 common variant sites between the two patients failed to reveal a familial relationship (Experimental Model and Subject Details). 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 ASEDA. The cohort used for this analysis was comprised mainly 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) 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 KIF5A are known causes of hereditary spastic paraparesis (spastic paraplegia type 10, autosomal dominant; OMIM: 604187) and of Charcot-Marie-Tooth disease type 2 (CMT2) (Crimella et al., 2012; 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 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 KIF5A. Interestingly, mutations contributing to SPG10 and to CMT2 are almost exclusively 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.

Patients with 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) (Ahmeti 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 (Ahmeti 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,

<table>
<thead>
<tr>
<th>Position</th>
<th>Variant</th>
<th>Exon</th>
<th>cDNA</th>
<th>Description</th>
<th>Predicted Exon Skipping</th>
<th>Gender</th>
<th>Age of Onset (Years)</th>
<th>Site of Onset</th>
<th>Survival (Months)</th>
<th>Alive (Yes/No)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control Variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>P</td>
<td>M</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</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>
<td>–</td>
<td>F</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</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>
<td>N</td>
<td>F</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>FALS Variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>–</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>27</td>
<td>c.2993-3C&gt;T</td>
<td>5’ splice junction</td>
<td>Y</td>
<td>M</td>
<td>29</td>
<td>L</td>
<td>&gt;264</td>
<td>Y</td>
</tr>
<tr>
<td>57,976,385</td>
<td>GA&gt;G</td>
<td>27</td>
<td>c.2996delA</td>
<td>p.Asn999fs</td>
<td>–</td>
<td>M</td>
<td>42</td>
<td>L</td>
<td>&gt;12</td>
<td>Y</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>
<td>Y</td>
<td>F</td>
<td>53</td>
<td>L</td>
<td>45</td>
<td>N</td>
</tr>
<tr>
<td>57,976,412</td>
<td>G&gt;A</td>
<td>27</td>
<td>c.3020G&gt;A</td>
<td>p.Arg1007Lys</td>
<td>Y</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>27</td>
<td>c.3020+1G&gt;A</td>
<td>3’ splice junction</td>
<td>Y</td>
<td>M</td>
<td>45</td>
<td>B</td>
<td>&gt;220</td>
<td>Y</td>
</tr>
<tr>
<td>57,976,414</td>
<td>T&gt;A</td>
<td>27</td>
<td>c.3020+2T&gt;A</td>
<td>3’ splice junction</td>
<td>Y</td>
<td>M</td>
<td>46</td>
<td>B</td>
<td>124</td>
<td>N</td>
</tr>
<tr>
<td>57,976,415</td>
<td>A&gt;G</td>
<td>27</td>
<td>c.3020+3A&gt;G</td>
<td>3’ splice junction</td>
<td>Y</td>
<td>M</td>
<td>50</td>
<td>B</td>
<td>54</td>
<td>N</td>
</tr>
<tr>
<td>SALS Variants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>N</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>57,975,731</td>
<td>CA&gt;C</td>
<td>26</td>
<td>c.2989delA</td>
<td>p.Asn997fs</td>
<td>–</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>27</td>
<td>c.2993-1G&gt;A</td>
<td>5’ splice junction</td>
<td>Y</td>
<td>n/a</td>
<td>52</td>
<td>B</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

P, possible; Y, yes; N, no; M, male; F, female; L, limb onset; B, bulbar onset; n/a, not available or applicable. Note that ASSEDA does not predict exon skipping based on frameshifts or nonsense mutations (Tompson et al., 2007).
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 based selected 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 GWASs, we cannot rule out false positives. In an independent line of investigation, we speculated that mutations in KIF5A contribute to motor neuron degeneration pathogenesis (Chevalier-Larsen and Holzbaur, 2006; Hirokawa et al., 2010; Millet-camps 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., 2004, 2005). 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 is a pathological hallmark of ALS and rare mutations in neurofilament heavy polypeptide (NEFH) are associated with ALS (Al-Chalabi et al., 1999).

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 mediate 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., 2004, 2005). 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 is 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 mitochondria (Kanai et al., 2014; Setou et al., 2002) and GABA 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 et al., 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 CMT2. 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 CMT2, 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 anterogradely 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-terminus 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 (Fares et al., 2017).

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 or CMT2) or more severe (i.e., ALS) phenotypes.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Study cohorts
- **METHOD DETAILS**
  - Data generation and pre-processing
  - Functional annotation of variants identified by WXS/WGS
  - RT-PCR Analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Statistical analyses
- **DATA AND SOFTWARE AVAILABILITY**
  - Datasets
  - Data Resources and Databases
  - Software

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes eight figures, eight tables, and consortia memberships and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.02.027.

**CONSORTIA**

The members of the ALS Sequencing Consortium are Andrew S. Allen, Stanley Appel, Robert H. Baloh, Richard S. Bedlack, Braden E. Boone, Robert Brown, John P. Carulli, Alessandra Chesi, Wendy K. Chung, Elizabeth T. Cirulli, Gregory M. Cooper, Julien Couthouis, Aaron G. Day-Williams, Patrick A. Dion, Summer Gibson, Aaron D. Gitter, Jonathan D. Glass, David B. Goldstein, Yujun Han, Matthew B. Harms, Tim Harris, Sebastian D. Hayes, Angela L. Jones, Jonathan Keebler, Brian J. Krueger, Brittany N. Lasseigne, Shawn E. Levy, Yi-Fan Lu, Tom Maniatis, Diane McKenna-Yasek, Timothy M. Miller, Richard M. Myers, Slavé Petrovski, Stefan M. Pulst, Alya R. Raphael, John M. Ravits,

The members of the ALS Foundation are Julia Kaye, Steven Fink-beiner, Stacia Wyman, Alexander LeNail, Leandro Lima, Ernst Freanekel, Jeffrey D. Rothstein, Clive N. Svendsen, Leslie M. Thompson, Jenny Van Eyk, Nicholas J. Maragakis, James D. Berry, Jonathan D. Glass, Timothy M. Miller, Stephen J. Kolb, Robert H. Bihol, Merit Kudowicz, and Emily Baxi.


The members of the ITALSGN Consortium are Francesco O. Logoli, Isabella Simone, Giancarlo Logroscino, Fabrizio Salvi, Ilaria Bartolomei, Giuseppe Bergoro, Claudia Rita Mimu, Emanuela Costantino, Carla Pani, Roberta Puddu, Carla Caredda, Valeria Piras, Stefania Tranquilli, Stefania Cuccu, Daniele Corongiu, Maurizio Melis, Antonio Milia, Francesca Moros, Maria Giovanna Marroso, Gisella Floris, Antonio Cannas, Stefania Tranquilli, Margherita Capasso, Claudia Caponnetto, Gianluigi Mancardi, Paola Orione, Stefania Tranquilli, M. theater, and GWendon Lemasson.

The members of the NYGC ALS Consortium are Hemali Phatnani, Justin Kwan, Dhruv Sareen, James R. Broach, Zachary Simmons, Ximena Arcila-Seppe Marrali, Giuseppe Fuda, Irene Ossola, Stefania Cammarosano, Antonio Michele Benigni, Tea B. Cau, Daniela Loi, Andrea Calvo, Cristina Moglia, Marialuisa Santarelli, Antonio Petrucci, Maura Pugliatti, Angelo Pirisi, Angela Pugliatti, Mauro Lanfi, Andrea Calvo, Cristina Moglia, Maura Brunetti, Mario Barberis, Gabriella Restagno, Federico Casale, Giuseppe Marrali, Giuseppe Fuda, Irene Ostella, Stefania Cammarosano, Antonio Casono, Antonio Iardi, Umberto Manera, Maurizio Grassano, Raffaella Tanel, and Fabrizio Pisanu.

The members of the NYGC ALS Consortium are Hemali Phatnani, Justin Kwan, Dhruv Sareen, James R. Broach, Zachary Simmons, Ximena Arcila-Seppe Marrali, Giuseppe Fuda, Irene Ossola, Stefania Cammarosano, Antonio Michele Benigni, Tea B. Cau, Daniela Loi, Andrea Calvo, Cristina Moglia, Maura Brunetti, Mario Barberis, Gabriella Restagno, Federico Casale, Giuseppe Marrali, Giuseppe Fuda, Irene Ostella, Stefania Cammarosano, Antonio Casono, Antonio Iardi, Umberto Manera, Maurizio Grassano, Raffaella Tanel, and Fabrizio Pisanu.

The members of the NYGC ALS Consortium are Hemali Phatnani, Justin Kwan, Dhruv Sareen, James R. Broach, Zachary Simmons, Ximena Arcila-Seppe Marrali, Giuseppe Fuda, Irene Ossola, Stefania Cammarosano, Antonio Michele Benigni, Tea B. Cau, Daniela Loi, Andrea Calvo, Cristina Moglia, Maura Brunetti, Mario Barberis, Gabriella Restagno, Federico Casale, Giuseppe Marrali, Giuseppe Fuda, Irene Ostella, Stefania Cammarosano, Antonio Casono, Antonio Iardi, Umberto Manera, Maurizio Grassano, Raffaella Tanel, and Fabrizio Pisanu.

The members of the NYGC ALS Consortium are Hemali Phatnani, Justin Kwan, Dhruv Sareen, James R. Broach, Zachary Simmons, Ximena Arcila-Seppe Marrali, Giuseppe Fuda, Irene Ossola, Stefania Cammarosano, Antonio Michele Benigni, Tea B. Cau, Daniela Loi, Andrea Calvo, Cristina Moglia, Maura Brunetti, Mario Barberis, Gabriella Restagno, Federico Casale, Giuseppe Marrali, Giuseppe Fuda, Irene Ostella, Stefania Cammarosano, Antonio Casono, Antonio Iardi, Umberto Manera, Maurizio Grassano, Raffaella Tanel, and Fabrizio Pisanu.
Helsinki University Hospital and the Sigrid Jusélius Foundation. J.D.G. received funding support from the ALS Association and Muscular Dystrophy Association. Additional funding was provided by the NIH/NINDS (R01NS073873, J.E.L.), the ALS Association (N.T., V.S., C.E.S., R.H.B.J., and J.E.L.), and the MND Association (N.T., V.S., C.E.S., and J.E.L.). J. Kaye, S.F., S.K.W., A.L., E.F., C.N.S., L.M.T., J.E.V.E., and J.D.R. received funding through NeurolINCS (NIH U54 NS091046). The sequencing activities at NYGC were additionally supported by the TOW Foundation. The CReAte consortium (U54NS092091) is part of the 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. The Target ALS Human Postmortem Tissue Core received funding support from Target ALS (grant 90072272). The InChIANTI study baseline (1998–2000) was supported as a “targeted project” (IC5110.1/RF97.71) by the Italian Ministry of Health and in part by the United States National Institute on Aging (contracts 263 MD 9164 and 263 MD 821336), the InChIANTI follow-up 1 (2001–2003) was funded by the United States National Institute on Aging (contract N1-A-1-1 and N1-A-1-211), and the InChIANTI follow-up 2 and 3 studies (2004–2010) were financed by the United States National Institute on Aging (contract N01-AG-5-0002). This work was supported in part by the Intramural Research Programs of the 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 Viali e Mauro Onlus, Compagnia di San Paolo, and European Community’s Health Seventh Framework Programme (FP7/2007-2013) under grant agreement 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; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. 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 of Aging (BLSA) and the Johns Hopkins University Alzheimer’s Disease Research Center, the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, and the Australian MND DNA Bank. We also thank Crystal Paccut, Blake Swihart, and Jayna Duell, RN for assistance in study coordination (S.A.G.). This study utilized the high-performance computational capabilities of the Biovulf Linux cluster at the NIH, Bethesda, Maryland (http://hpc.nih.gov), and the Massachusetts Green High Performance Computing Center at the University of Massachusetts Medical School (https://www.mghcc.org). This study also used genotype and clinical data from the Welcome Trust Case Control Consortium.

AUTHOR CONTRIBUTIONS

DECLARATION OF INTERESTS
J.D.B. is a consultant to Neuraltus Pharmaceuticals and Denali Therapeutics, and held a research fellow position funded by Voyager Therapeutics. M. Cudkowicz 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 for 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...


Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet Neurol. 9, 986–994.


## STAR★METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIzol Reagent</td>
<td>Thermo Fisher Scientific</td>
<td>Cat# 15596026</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><strong>Critical Commercial Assays</strong></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>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 primer sequence: CAGTGGAGCCACATCTTCTG</td>
<td>Operon Technologies</td>
<td>N/A</td>
</tr>
<tr>
<td>R1 primer sequence: TCTCTTGGTGGAGAGGGAAA</td>
<td>Operon Technologies</td>
<td>N/A</td>
</tr>
<tr>
<td>F2 primer sequence: CCAACATGGACAATGGAGTGA</td>
<td>Operon Technologies</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Software and Algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASSEDA</td>
<td>Tompson et al., 2007</td>
<td><a href="http://www.cytognomix.com/?post_type=duka&amp;p=2670">http://www.cytognomix.com/?post_type=duka&amp;p=2670</a></td>
</tr>
<tr>
<td>BWA</td>
<td>Wellcome Trust Sanger Institute</td>
<td><a href="http://bio-bwa.sourceforge.net">http://bio-bwa.sourceforge.net</a></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_genom_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>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><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 Tables S1 and 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. (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, 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 (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 SNP. 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 μg with RNase inhibitor in a 20 μl reaction according to the manufacturer’s protocol. PCR was carried out using New England Biolabs One Taq Hot Start DNA Polymerase (# M0481S), 2 μl RT reaction (representing 50 ng input RNA) and forward and reverse primer (0.15 μM each) in a 20 μl reaction volume. Amplification conditions were as follows: 94°C for 30 s, (94°C for 20 s, 58°C for 20 s, 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 (TCTCTTGTGGAAGGAAA). Primers used for the specific amplification of the mutant splice form were F2 (CCAACATGGACAATGGAGTGA), 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, http://www.internationalgenome.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 as co-variables 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 R^2 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 (https://github.com/AudeDN/ALS_GWAS_1000G_mach2dat_2017), and GWAS summary statistics results for all tested SNPs are available from https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000101.v4.p1.

**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, heterozygosity, 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 and 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 minimum $p < 1 \times 10^{-3}$ 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. The accession number for the genotyped data reported in this paper is dbGaP: phs000101.v4 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000101.v4.p1).

**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: https://cordis.europa.eu/project/rcn/86758_en.html
Project MinE Variant Browser: http://databrowser.projectmine.com
snpEFF: http://snpeff.sourceforge.net/SnpEff.html
Wellcome Trust Case Control Consortium: https://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