Genome-wide Analyses Identify KIF5A as a Novel ALS Gene

(Article begins on next page)

Genome-wide Analyses Identify KIF5A as a Novel ALS Gene

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 spectrum of subphenotypes 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., 2017; 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 processing (based on the observation of mutations in C9orf72, TDP-43, FUS, and MATR3); (2) protein homeostasis (UBQLN2, VCP, OPTN, VAPB); (3) cytoskeletal dynamics (PFN1, TUBA4A, DCTN1)(Robberecht and Eykens, 2015). 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 identification of ALS associated genes with moderate to low impact. Despite their lower impact, such genes continue to provide valuable insight into ALS pathogenesis. For example, 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 of index FALS, is a known binding partner of C21orf2, an ALS risk factor found through 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 recognized 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).

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

To identify new susceptibility loci operating in ALS, we undertook a large-scale genome-wide association study (GWAS) involving 12,663 patients diagnosed with ALS and 53,439 control subjects (**Table S1, S2**). Our data was 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 (**Methods, Figure S1**), 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 S2**). 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 genes that achieved genome-wide significance (*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 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*=6.4x10^{-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 5/529/12,043:7/786/22,682 (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 (**Figure S3**).

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 (**Methods, Figure S4, Table S5)**. Genes displaying $P< 5 \times 10^{-4}$ are shown in **Table 2**. The previously identified ALS genes, *TBK1* (P=5.58 x 10^{-7}, OR= 15.11, 95%CI = 5.81-38.69) and *NEK1* (P=1.68 x 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=5.55x10^{-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 (λ =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.

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 (Sheth et al., 2006). This algorithm 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 confirmed by performing RT-PCR using RNA from lymphoblasts available for a single patient carrying the c.3020+2T>A, exon 27 - 3' splice junction variant and two control lines (**Figure 4D**). Lymphoblasts were 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 3.8×10^{-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 S7**).

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 (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 4,160 ALS cases and 18,650 controls that were non-overlapping with our GWAS discovery analysis, after applying quality control filters (**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 1.73% in cases and 1.32% in controls (rs113247976, $P=1.24 \times 10^{-3}$, OR=1.38, 95% CI=1.14-1.66), thereby replicating the association of the original GWAS. A meta-analysis of the GWAS and replication cohort (n=24,966 cases, 78,454 controls) yielded a highly significant P of 2.31×10^{-12} (OR=1.38, 95%CI, 1.26-1.51)(**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.

Additional Screening of LOF Variants in *KIF5A*

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 two additional carriers of C-terminal variants, namely a carrier of an exon 26 frameshift mutation (p.Asn997fs) and a carrier of an exon 27 splice altering mutation (c.2993-

1G>A, **Table 3**). 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 frequency of LOF variants in this cohort (2/9,046 cases, 0.022%), which primarily comprised of sporadic ALS cases, is lower than that observed in our original FALS cohort (0.703%), suggesting that *KIF5A* LOF variants display a high penetrance. LOF variants were not observed in a follow up panel of 1,955 controls.

ALS-Associated Mutations in *KIF5A* are Distinct from SPG10/CMT2 Mutations

Missense mutations within *KIF5A* are a known cause of hereditary spastic paraparesis (spastic paraplegia type 10, autosomal dominan; 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 *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 found almost exclusively missense changes and 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 S6**). 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 S6**). 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 previous 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). Here, we have

confirmed *KIF5A* as an ALS-associated gene through two independent approaches. By performing a GWAS involving ~80,000 samples, we identified a missense variant within the *KIF5A* gene that reached genome-wide significance for association with ALS risk (P=6.4 x 10^{-10}, OR= 1.38). In an independent line of investigation, we applied RVB to exome sequencing of ~21,000 samples and identified an exome-wide significant association between FALS risk and rare *KIF5A* LOF variants (OR=41.16, P=3.8 x 10^{-9}). The GWAS and the exome gene burden signals were completely independent, as none of the cases carrying the p.Pro986Leu variant also carry LOF mutations. Follow up analyses of *KIF5A* in independent replication cohorts confirmed our initial finding and resulting in a highly significant association for both the p.Pro986Leu variant resulting (P=2.37 x 10^{-12}, OR=1.38) and revealed two 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). The complex of KIF5s and kinesin light chains is called Kinesin-1. 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 lead 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 (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; 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 domain is associated with ALS, while variation of the N-terminus is associated with hereditary spastic paraparesis and Charcot-Marie-Tooth, type 2. KIF5A can be structurally divided into three domains: of these, the N-terminal domain binds microtubules, and is responsible for ATP hydrolysis and the kinesin motor activity; the central coiled-coil or stalk domain mediates heavy chain dimerization, while the C-terminal or tail domain interacts with cargo molecules. Notably, both the p.Pro986Leu and ALS-associated LOF variants are consistently localized within the C-terminal domain of KIF5A. Conversely, missense mutations previously associated with hereditary spastic paraparesis and Charcot-Marie-Tooth, type 2 have been found in the central or N-terminal motor domain. Missense mutations within this latter domain 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). Conversely, 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. Interestingly, LOF variants within the C-terminal domain of KIF5A are

expected to disrupt binding with specific cargo proteins, possibly leading to their

accumulation and seeding aggregation within the cell body and their deficiency at neurite

terminals. Deficiency in KIF5A expression and cargo binding has been associated to

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 that 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.

EXPERIMENTAL PROCEDURES

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 dbGaP web

repository (www.ncbi.nlm.nih.gov/gap, n = 44,017 US samples) analysis; (b) the

HYPERGENES Project (n = 887 Italian samples) (Salvi et al., 2012); and (c) the

Wellcome Trust Case Control Consortium (www.wtccc.org.uk, 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 online in the Supplemental Information.

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

databrowser.projectmine.com. Additional details of the cohorts used in this study are

available in van Rheenen *et al* (van Rheenen et al., 2016).

FALS WXS discovery cohort. A total of 1,463 FALS patients were subjected to analysis

by exome sequencing. 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), USA (n=417). Familial history was considered positive for ALS if the proband had at least one affected relative within three degrees of relatedness.

Control WXS discovery cohort. Read level exome sequencing data was obtained for a total of 41,410 non-ALS controls deposited in dbGAP and the EGA by various projects. Utilized dbGaP datasets include the Alzheimer's Disease Sequencing Project (ADSP) phs000572; University of Miami Udall Center of Excellence Identification of Rare Variants in PD through Whole Exome Sequencing phs000908; Myocardial Infarction Genetics Exome Sequencing Consortium: Ottawa Heart Study phs000806; Myocardial Infarction Genetics Exome Sequencing Consortium: Italian Atherosclerosis Thrombosis and Vascular Biology phs000814; Myocardial Infarction Genetics Exome Sequencing Consortium: Pakistan Risk Of Myocardial Infarction Study phs000917; Myocardial Infarction Genetics Exome Sequencing Consortium: U. of Leicester phs001000; Myocardial Infarction Genetics Exome Sequencing Consortium: Malmo Diet and Cancer Study phs001101; Yale Center for Mendelian Genomics (Y CMG) phs000744; Framingham Heart Study Allelic Spectrum Project phs000307; NHLBI GO-ESP Lung Cohorts Exome Sequencing Project: Genetic modifiers of Pseudomonas aeruginosa (Pa) lung infection acquisition in cystic fibrosis phs000254; National Heart Lung and Blood Institute (NHLBI) GO-ESP: Heart Cohorts Component of the Exome Sequencing Project (FHS) phs000401; National Heart Lung and Blood

Institute (NHLBI) GO-ESP: Heart Cohorts Component of the Exome Sequencing Project (JHS) phs000402; National Heart Lung and Blood Institute (NHLBI) GO-ESP: Heart Cohorts Component of the Exome Sequencing Project (ARIC) phs000398; NHLBI GO-ESP: Women's Health Initiative Exome Sequencing Project (WHI) - WHISP phs000281; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Pulmonary Arterial Hypertension) phs000290; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Lung Health Study of Chronic Obstructive Pulmonary Disease) phs000291; National Heart Lung and Blood Institute (NHLBI) GO-ESP: Heart Cohorts Component of the Exome Sequencing Project (CHS) phs000400; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (COPDGene) phs000296; JHS Allelic Spectrum Project phs000498; NHLBI GO-ESP Family Studies: Idiopathic Bronchiectasis of unknown etiology that is not related to cystic fibrosis or classic primary ciliary dyskinesia or immune deficiency or any other known causes phs000518; Genetic Epidemiology of COPD (COPDGene) Funded by the National Heart, Lung, and Blood Institute phs000179; National Heart Lung and Blood Institute (NHLBI) GO-ESP: Heart Cohorts Component of the Exome Sequencing Project (MESA) phs000403; NHLBI GO-ESP: Family Studies (Thoracic aortic aneurysms leading to acute aortic dissections) phs000347; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Asthma): Genetic variants affecting susceptibility and severity phs000422; NHLBI GO-ESP: Family Studies (Familial Atrial Fibrillation) phs000362; NHLBI GO-ESP Family Studies: Pulmonary Arterial Hypertension phs000354; Building on GWAS: the U.S. CHARGE consortium - Sequencing (CHARGE-S): FHS phs000651; NHLBI GO-ESP: Family Studies (Hematological Cancers) phs000632; Building on GWAS for NHLBI-Diseases : the U.S.

CHARGE Consortium (CHARGE-S): ARIC phs000668; Building on GWAS for

NHLBI-Diseases: The U.S. CHARGE Consortium (CHARGE-S): CHS phs000667.

Utilized EGA projects included the UK10K COHORT ALSPAC EGAS00001000090;

UK10K COHORT TWINSUK EGAS00001000108; UK10K NEURO ABERDEEN

EGAS00001000109; UK10K NEURO ASD GALLAGHER EGAS00001000112;

UK10K NEURO EDINBURGH EGAS00001000117; UK10K NEURO IOP COLLIER

EGAS00001000121; UK10K_NEURO_MUIR EGAS00001000122;

UK10K_RARE_HYPERCHOL EGAS00001000129; UK10K_RARE_SIR

EGAS00001000130; UK10K_RARE_THYROID EGAS00001000131;

UK10K_OBESITY_GS EGAS00001000242.

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 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; 161 cases

subjected to WGS by the Create Consortium; 1,017 cases subjected to WGS by the

Centre for genomics of neurodegenerative diseases; 4,100 cases and 1,659 controls

subjected to WGS by the project MinE 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).

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 (www.ncbi.nlm.nih.gov/gap, 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 best

practices (ref). 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). All identified LOF variants were validated by Sanger sequencing.

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. All identified LOF variants were validated by Sanger sequencing.

Functional annotation of variants identified by WXS/WGS

Variant calls were assigned predicted functional consequences using snpEFF (Single Nucleotide Polymorphism Effect) and dbscSNV (database of splice site consequences of Single Nucleotide Variants). 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, 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)(Mucaki et al., 2013).

Statistical analyses.

Analysis of SNP array genotypes. Standard quality-control procedures were applied to our genotype data using PLINK software package (version 1.9, PMID: 25722852). 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.v2), 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 5.0×10^{-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, PMID: 23269371). 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, PMID: 27571263) 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, PMID: 20517342) 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, PMID:20616382) under an inverse-weighted, fixed effect model. A threshold *P* of 5.0×10^{-8} was set for genome-wide significance after Bonferroni correction for multiple testing (Pe'er et al., 2008).

The programming code used to analyze these data is freely available on GitHub (link to be supplied at publication), 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, heterozygosity, duplication, relatedness or population stratification filters as summarized in Table S5 and as previously described (Kenna et al., 2016). Each of these filters was performed using a set of autosomal markers meeting all of the following criteria: call rate>0.95, MAF>0.01, p>0.001 for deviation from Hardy Weinberg equilibrium, LD 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 v1.09. Study duplicates and sample relatedness within the WXS/WGS cohorts was identified using KING. Study duplicates between WXS/WGS cohorts and GWAS datasets were identified using the checksum program (id_geno_checksum.v2, https://personal.broadinstitute.org/sripke/share_links/checksums_download/). Principal components analyses were performed using GCTA (Genome-wide Complex Trait Analysis).

RVB analyses were performed by penalized logistic regression of case-control status with respect to number of minor alleles observed per sample per gene as described previously (Kenna et al., 2016). Analyses were only performed where the dataset contained >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 (GenABEL).

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 v3.2.0.

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 using manufacturer's methods. PCR was carried out using NEB 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. Amplification conditions were as follows: 94°C 30 sec, {94°C 20 sec, 58°C 20 sec, 68°C 1 min} x 35, 68°C 5 min, 4°C hold. Amplification of both normal and mutant splice forms used primers F1 (CAGTGGAGCCACATCTTCTG) and R1 (TCTCTTGGTGGAGAGGGAAA). Primers used for the specific amplification of the mutant splice form were F2 (CCAACATGGACAATGGAGTGA), which spans exons 26 and 28, and R1.
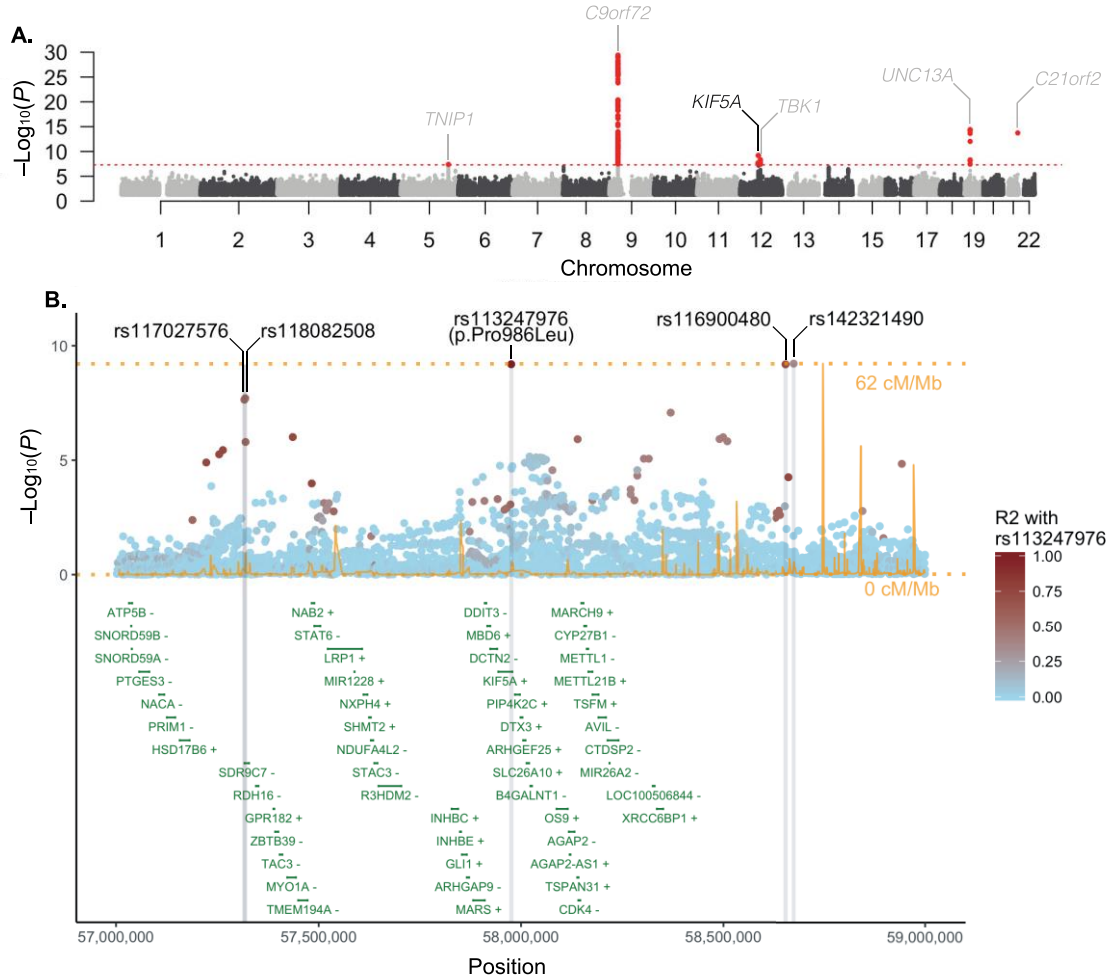
Figure 1. Identification of association between *KIF5A* locus and ALS risk through GWAS. (A) Manhattan plot showing p-values from the discovery set GWAS. Analysis of a combined set of 20,806 cases and 59,804 controls is shown. The dashed red line denotes the threshold for genome-wide significance after multiple test correction ($P < 5.0 \times 10^{-8}$). Five previously reported ALS associated genes are labeled in grey and one novel gene, *KIF5A*, 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 most significant SNP per locus 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).

Cohort	Cases	Controls	Case MAF	Control MAF	OR (95% CI)	P	
Italy	2,853	2,143	1.28%	0.82%	1.45 (0.90-2.34)		
United Kingdom	449	226	3.56%	1.33%	2.07 (0.85-5.07)		
Belgium / France	1,150	595	1.91%	1.26%	1.42 (0.75-2.68)		
USA	3,777	33,365	2.01%	1.46%	1.45 (1.19-1.75)		
Present Study	**8,229**	**36,329**	**1.83%**	**1.42%**	**1.46 (1.24-1.73)**	9.23×10^{-6}	
Van Rheenen Study	12,577	23,475	2.23%	1.75%	1.33 (1.17-1.52)	1.13×10^{-5}	
Combined Discovery Set	**20,806**	**59,804**	**2.07%**	**1.55%**	**1.38 (1.25-1.53)**	6.43×10^{-10}	
Replication	4,160	18,650	1.73%	1.32%	**1.38 (1.14-1.66)**	1.24×10^{-3}	
Joint	24,966	78,454	2.01%	1.49%	**1.38 (1.26-1.51)**	2.31×10^{-12}	

0.70   1.0   1.38   2.0   3.0

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 cohorts described is presented. 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 vertical dotted line denoting OR estimated under the meta-analysis.
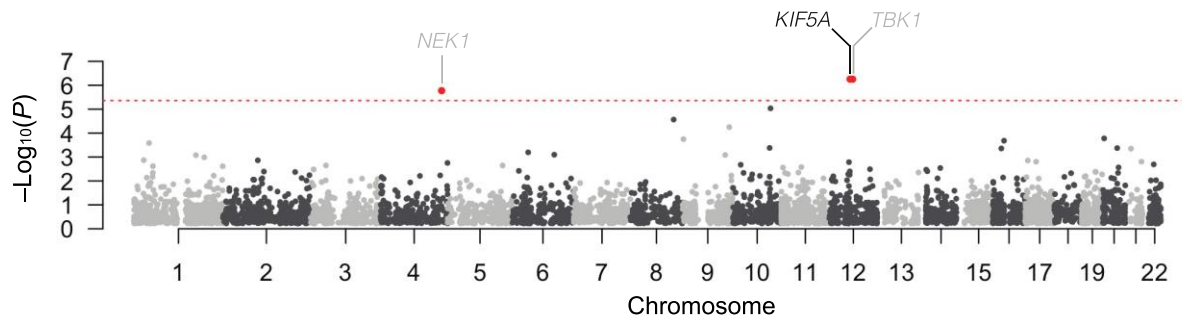
Figure 3. Identification of association between *KIF5A* and ALS risk through rare variant burden analysis of exome sequencing. Manhattan plot showing gene-level p-values from an exome-wide rare variant burden analysis. Analyses of 1,138 index FALS cases vs. 19,494 controls were restricted to rare LOF variants as previously defined (splice altering/nonsense, MAF<0.001)(Kenna et al., 2016). 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 (4.36 x 10^{-6}). Previously reported (grey) and novel (black) genes exhibiting a significant excess of rare LOF variants in patients are shown.
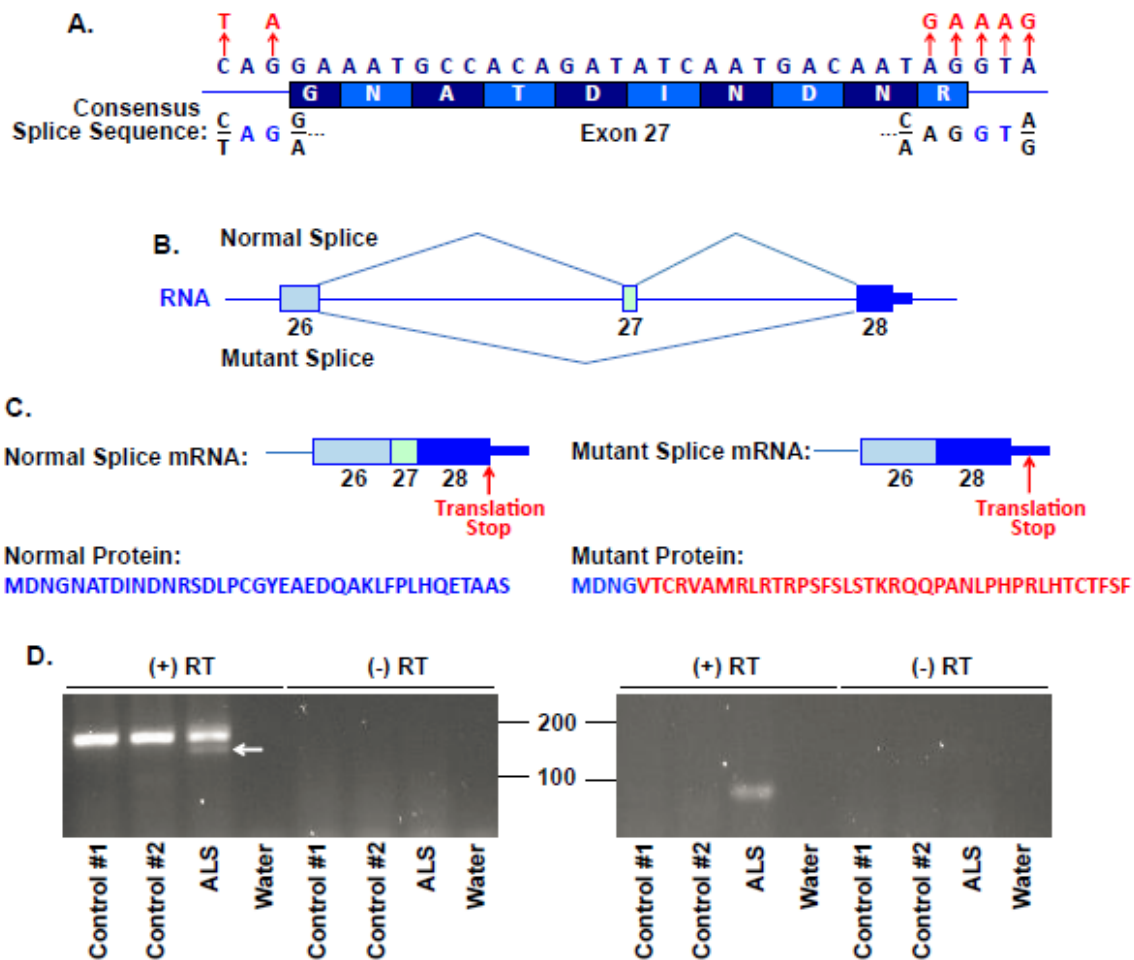
Figure 4. ALS associated loss of function variants of KIF5A disrupt C-terminal sequence by inducing skipping of exon 27. (A) Single nucleotide variants (SNVs) within *KIF5A* identified in ALS patients are clustered at the 5' and 3' splice junctions of exon 27. The consensus splice sequence is shown. (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 from lymphoblasts derived from ALS patients without (Control) or with (ALS) the loss of function c.3020+2T>A allele predicted to induce skipping of exon 27. PCR reactions used primers to either amplify both normal (155 bp) and mutant (127 bp) splice forms (left panel) or specifically the mutant splice form (80 bp)(right). The arrow represents the position of the mutant specific product in the left panel.

SPG10/CMT2
p.Y63C, p.D73N, p.R162W, p.M198T, p.S202N, p.S203C, p.R204Q, p.R204W, p.V231L, p.D232N, p.G235E, p.E251K, p.K253N, p.K256del, p.N256S, p.K257N, p.S258L, p.L259Q, p.Y276C, p.P278L, p.R280H, p.R280C, p.R280L, p.R323W, p.A361V, p.E755K

ALS
p.Pro986Leu**, c.2993-3C>T, p.Arg1007Gly, p.Arg1007Lys, c.3020+1G>A, c.3020+2T>A, c.3020+3A>G, p.Asp996fs, p.Asn999fs, p.Asn997fs, c.2993-1G>A, p.Asn999del

Motor Domain
Microtubule Binding, Kinesin Motor
(9-327)

Stalk
Heavy Chain Dimerization
(331-906)

Tail
Cargo Binding
(907-1032)

Figure 5. KIF5A ALS mutations show distinct localization to 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 mutation identified within this study are shown. As illustrated, mutations causative for SPG10/CMT2 are predominantly missense changes and located in the N-terminal motor domain. In contrast, ALS mutations are primarily located at the C-terminal motor domain with most penetrant displaying LOF.

SNP Information				Present Study (8,229 Cases / 36,329 Controls)				Van Rheenen *et al.* (12,577 Cases / 23,475 Controls)				Combined Discovery Set (20,806 Cases / 59,804 Controls)			
SNP	Chr	Position	Gene	Case MAF	Control MAF	OR [95% CI]	*P*	Case MAF	Control MAF	OR [95% CI]	*P*	Case MAF	Control MAF	OR [95% CI]	*P*
Novel Loci															
rs117027576	12	57,316,603	*KIF5A*	1.55%	1.27%	1.45 [1.20-1.76]	1.1×10^{-4}	1.98%	1.59%	1.33 [1.16-1.53]	4.3×10^{-5}	1.81%	1.40%	1.37 [1.23-1.54]	2.3×10^{-8}
rs118082508	12	57,318,819	*KIF5A*	1.56%	1.28%	1.45 [1.20-1.76]	1.0×10^{-4}	1.98%	1.60%	1.33 [1.16-1.53]	3.8×10^{-5}	1.81%	1.41%	1.37 [1.23-1.54]	2.0×10^{-8}
rs113247976*	12	57,975,700	*KIF5A*	1.83%	1.42%	1.46 [1.23-1.74]	9.2×10^{-6}	2.14%	1.70%	1.33 [1.17-1.52]	1.1×10^{-5}	2.02%	1.53%	1.38 [1.24-1.53]	6.4×10^{-10}
rs116900480	12	58,656,105	*KIF5A*	1.75%	1.46%	1.42 [1.21-1.68]	1.9×10^{-5}	2.08%	1.66%	1.34 [1.18-1.53]	7.1×10^{-6}	1.95%	1.54%	1.37 [1.24-1.52]	6.6×10^{-10}
rs142321490	12	58,676,132	*KIF5A*	1.79%	1.48%	1.43 [1.21-1.68]	1.5×10^{-5}	2.08%	1.66%	1.34 [1.18-1.53]	8.0×10^{-6}	1.97%	1.55%	1.37 [1.24-1.52]	6.1×10^{-10}
Previously Published Loci															
rs10463311	5	150,410,835	*TNIP1*	73.19%	74.84%	0.94 [0.89-0.98]	7.8×10^{-3}	73.34%	75.79%	0.91 [0.87-0.94]	8.5×10^{-7}	73.28%	75.21%	0.92 [0.89-0.95]	4.0×10^{-8}
rs3849943	9	27,543,382	*C9orf72*	71.79%	76.31%	0.84 [0.80-0.88]	1.4×10^{-12}	72.78%	76.5%	0.83 [0.80-0.87]	4.0×10^{-19}	72.39%	76.38%	0.84 [0.81-0.86]	3.8×10^{-30}
rs74654358	12	64,881,967	*TBK1*	3.77%	4.01%	1.20 [1.07-1.34]	1.6×10^{-3}	5.12%	4.61%	1.23 [1.13-1.34]	7.7×10^{-7}	4.59%	4.25%	1.22 [1.14-1.30]	4.7×10^{-9}
rs12973192	19	17,753,239	*UNC13A*	67.62%	69.37%	0.86 [0.82-0.91]	1.3×10^{-8}	64.52%	66.00%	0.9 [0.87-0.93]	2.4×10^{-8}	65.75%	68.05%	0.89 [0.86-0.91]	3.9×10^{-15}
rs75087725	21	45,753,117	*C21orf2*	0.70%	0.46%	1.99 [1.44-2.75]	2.2×10^{-5}	1.83%	1.27%	1.61 [1.39-1.87]	8.7×10^{-11}	1.38%	0.78%	1.67 [1.46-1.91]	1.8×10^{-14}

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).

Gene	FALS	Control	OR	P
KIF5A	6 (0.53%)	3 (0.02%)	32.07 (9.05-135.27)	5.55×10^{-7}
TBK1	8 (0.70%)	9 (0.05%)	15.11 (5.81-38.69)	5.58×10^{-7}
NEK1	12 (1.05%)	32 (0.16%)	6.64 (3.32-12.51)	1.68×10^{-6}
CALHM2	7 (0.62%)	9 (0.05%)	12.13 (4.47-31.79)	9.19×10^{-6}
COL14A1	8 (0.70%)	16 (0.08%)	8.04 (3.32-18.08)	2.72×10^{-5}
AK1	10 (0.88%)	34 (0.17%)	5.37 (2.55-10.41)	5.62×10^{-5}
ATRN	5 (0.44%)	9 (0.05%)	11.06 (3.57-31.02)	1.66×10^{-4}
VLDLR	5 (0.44%)	9 (0.05%)	10.87 (3.51-30.43)	1.79×10^{-4}
FUS	4 (0.35%)	4 (0.02%)	16.53 (4.25-64.33)	2.08×10^{-4}
ZMYND12	6 (0.53%)	12 (0.06%)	7.92 (2.86-19.96)	2.61×10^{-4}

Table 2. Top ALS associations identified through RVB of FALS and control exome sequencing results.

Position	Variant	Exon	cDNA	Description	Predicted Exon Skipping	Gender	Age of Onset (years)	Site of Onset	Survival (months)	Alive (yes/no)
Control Variants										
57,963,470	A>G	11	c.1117+4A>G	3' Splice Junction	P	M	n/a	n/a	n/a	n/a
57,966,423	C>T	15	c.1630C>T	p.Arg544*	-	F	n/a	n/a	n/a	n/a
57,976,884	G>C	28	c.3021G>C	5' Splice Junction	N	F	n/a	n/a	n/a	n/a
FALS Variants										
57,975,729	GA>A	26	c.2987delA	p.Asp996fs	-	M	45	n/a	n/a	n/a
57,976,382	C>T	27	c.2993-3C>T	5' Splice Junction	Y	M	29	L	>264	Y
57,976,385	GA>G	27	c.2996delA	p.Asn999fs	-	M	42	L	>12	Y
57,976,411	A>G	27	c.3019A>G	p.Arg1007Gly	Y	F	53	L	45	N
57,976,412	G>A	27	c.3020G>A	p.Arg1007Lys	Y	M	50	L	>108	Y
57,976,413	G>A	27	c.3020+1G>A	3' Splice Junction	Y	M	45	B	>220	Y
57,976,414	T>A	27	c.3020+2T>A	3' Splice Junction	Y	M	46	B	124	N
57,976,415	A>G	27	c.3020+3A>G	3' Splice Junction	Y	M	50	B	54	N
SALS Variants										
57,957,481	G>A	3	c.291+5G>A	3' Splice Junction	N	n/a	n/a	n/a	n/a	n/a
57,975,731	CA>C	26	c.2989delA	p.Asn997fs	-	F	50	L	>96	Y
57,976,384	G>A	27	c.2993-1G>A	5' Splice Junction	Y	n/a	52	B	n/a	n/a

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 non-sense mutations (Mucaki et al., 2013).

REFERENCES

Al-Chalabi, A., Andersen, P.M., Nilsson, P., Chioza, B., Andersson, J.L., Russ, C., Shaw, C.E., Powell, J.F., and Leigh, P.N. (1999). Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. Human Molecular Genetics *8*, 157–164.

ALSGEN Consortium, Ahmeti, K.B., Ajroud-Driss, S., Al-Chalabi, A., Andersen, P.M., Armstrong, J., Birve, A., Blauw, H.M., Brown, R.H., Bruijn, L., et al. (2013). Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. Neurobiol. Aging *34*, 357.e7–.e19.

Arthur, K.C., Calvo, A., Price, T.R., Geiger, J.T., Chio, A., and Traynor, B.J. (2016). Projected increase in amyotrophic lateral sclerosis from 2015 to 2040. Nat Commun *7*, 12408.

Benyamin, B., He, J., Zhao, Q., Gratten, J., Garton, F., Leo, P.J., Liu, Z., Mangelsdorf, M., Al-Chalabi, A., Anderson, L., et al. (2017). Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. Nat Commun *8*, 611.

Brooks, B.R. (1994). El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. J. Neurol. Sci. *124 Suppl*, 96–107.

Chevalier-Larsen, E., and Holzbaur, E.L.F. (2006). Axonal transport and neurodegenerative disease. Biochim. Biophys. Acta *1762*, 1094–1108.

Chia, R., Chio, A., and Traynor, B.J. (2017). Novel genes associated with amyotrophic lateral sclerosis: diagnostic and clinical implications. Lancet Neurol.

Cirulli, E.T., Lasseigne, B.N., Petrovski, S., Sapp, P.C., Dion, P.A., Leblond, C.S., Couthouis, J., Lu, Y.-F., Wang, Q., Krueger, B.J., et al. (2015). Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science *347*, 1436–1441.

Coppedè, F., and Migliore, L. (2015). Mutation Research/Fundamental and MolecularMechanisms of Mutagenesis. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis *776*, 84–97.

Crimella, C., Baschirotto, C., Arnoldi, A., Tonelli, A., Tenderini, E., Airoldi, G., Martinuzzi, A., Trabacca, A., Losito, L., Scarlato, M., et al. (2011). Mutations in the motor and stalk domains of KIF5A in spastic paraplegia type 10 and in axonal Charcot-Marie-Tooth type 2. Clin. Genet. *82*, 157–164.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., et al. (2011).

Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245–256.

Ebbing, B., Mann, K., Starosta, A., Jaud, J., Schols, L., Schule, R., and Woehlke, G. (2008). Effect of spastic paraplegia mutations in KIF5A kinesin on transport activity. Human Molecular Genetics *17*, 1245–1252.

Fang, H., Wu, Y., Narzisi, G., O'Rawe, J.A., Barrón, L.T.J., Rosenbaum, J., Ronemus, M., Iossifov, I., Schatz, M.C., and Lyon, G.J. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med *6*, 89.

Fang, X., Lin, H., Wang, X., Zuo, Q., Qin, J., and Zhang, P. (2015). The NEK1 interactor, C21ORF2, is required for efficient DNA damage repair. Acta Biochim Biophys Sin *47*, 834–841.

Freischmidt, A., Wieland, T., Richter, B., Ruf, W., Schaeffer, V., Müller, K., Marroquin, N., Nordin, F., Hübers, A., Weydt, P., et al. (2015). Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. Nat Neurosci *18*, 631–636.

Gladman, M., and Zinman, L. (2015). The economic impact of amyotrophic lateral sclerosis: a systematic review. Expert Rev Pharmacoecon Outcomes Res *15*, 439–450.

Gros-Louis, F. (2004). A Frameshift Deletion in Peripherin Gene Associated with Amyotrophic Lateral Sclerosis. Journal of Biological Chemistry *279*, 45951–45956.

Hares, K., Redondo, J., Kemp, K., Rice, C., Scolding, N., and Wilkins, A. (2016). Axonal motor protein KIF5A and associated cargo deficits in multiple sclerosis lesional and normal-appearing white matter. Neuropathol Appl Neurobiol *43*, 227–241.

Heisler, F.F., Lee, H.K., Gromova, K.V., Pechmann, Y., Schurek, B., Ruschkies, L., Schroeder, M., Schweizer, M., and Kneussel, M. (2014). GRIP1 interlinks N-cadherin and AMPA receptors at vesicles to promote combined cargo transport into dendrites. Proceedings of the National Academy of Sciences *111*, 5030–5035.

Hirokawa, N., Pfister, K.K., Yorifuji, H., Wagner, M.C., Brady, S.T., and Bloom, G.S. (1989). Submolecular domains of bovine brain kinesin identified by electron microscopy and monoclonal antibody decoration. Cell *56*, 867–878.

Hirokawa, N., Niwa, S., and Tanaka, Y. (2010). Molecular Motors in Neurons:Transport Mechanisms and Rolesin Brain Function, Development, and Disease. Neuron *68*, 610–638.

Hirtz, D., Thurman, D.J., Gwinn-Hardy, K., Mohamed, M., Chaudhuri, A.R., and Zalutsky, R. (2007). How common are the "common" neurologic disorders? Neurology *68*, 326–337.

Jennings, S., Chenevert, M., Liu, L., Mottamal, M., Wojcik, E.J., and Huckaba, T.M. (2017). Characterization of kinesin switch I mutations that cause hereditary spastic

paraplegia. PLoS ONE *12*, e0180353.

Kanai, Y., Okada, Y., Tanaka, Y., Harada, A., Terada, S., and Hirokawa, N. (2000). KIF5C, a novel neuronal kinesin enriched in motor neurons. J. Neurosci. *20*, 6374–6384.

Kanai, Y., Dohmae, N., and Hirokawa, N. (2004). Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron *43*, 513–525.

Karle, K.N., Möckel, D., Reid, E., and Schols, L. (2012). Axonal transport deficit in a KIF5A –/– mouse model. Neurogenetics *13*, 169–179.

Kenna, K.P., van Doormaal, P.T.C., Dekker, A.M., Ticozzi, N., Kenna, B.J., Diekstra, F.P., van Rheenen, W., van Eijk, K.R., Jones, A.R., Keagle, P., et al. (2016). NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat. Genet. *48*, 1037–1042.

Kim, H.J., Kim, N.C., Wang, Y.-D., Scarborough, E.A., Moore, J., Diaz, Z., MacLea, K.S., Freibaum, B., Li, S., Molliex, A., et al. (2013). Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. Nature *495*, 467–473.

Kwiatkowski, T.J., Bosco, D.A., LeClerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., et al. (2009). Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. Science *323*, 1205–1208.

Laaksovirta, H., Peuralinna, T., Schymick, J.C., Scholz, S.W., Lai, S.-L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D.G., Gibbs, J.R., et al. (2010). Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol *9*, 978–985.

Liu, Y.-T., Laurá, M., Hersheson, J., Horga, A., Jaunmuktane, Z., Brandner, S., Pittman, A., Hughes, D., Polke, J.M., Sweeney, M.G., et al. (2014). Extended phenotypic spectrum of KIF5A mutations: From spastic paraplegia to axonal neuropathy. Neurology *83*, 612–619.

Lopez-Gonzalez, R., Lu, Y., Gendron, T.F., Karydas, A., Tran, H., Yang, D., Petrucelli, L., Miller, B.L., Almeida, S., and Gao, F.-B. (2016). Poly(GR) in C9ORF72-Related ALS/FTD Compromises Mitochondrial Function and Increases Oxidative Stress and DNA Damage in iPSC-Derived Motor Neurons. Neuron *92*, 383–391.

Madabhushi, R., Pan, L., and Tsai, L.-H. (2014). DNA Damage and Its Links to Neurodegeneration. Neuron *83*, 266–282.

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

Malovannaya, A., Lanz, R.B., Jung, S.Y., Bulynko, Y., Le, N.T., Chan, D.W., Ding, C., Shi, Y., Yucer, N., Krenciute, G., et al. (2011). Analysis of the human endogenous coregulator complexome. Cell *145*, 787–799.

Maruyama, H., Morino, H., Ito, H., Izumi, Y., Kato, H., Watanabe, Y., Kinoshita, Y., Kamada, M., Nodera, H., Suzuki, H., et al. (2010). Mutations of optineurin in amyotrophic lateral sclerosis. Nature *465*, 223–226.

Matsuzaki, F., Shirane, M., Matsumoto, M., and Nakayama, K.I. (2011). Protrudin serves as an adaptor molecule that connects KIF5 and its cargoes in vesicular transport during process formation. Molecular Biology of the Cell *22*, 4602–4620.

McLaughlin, R.L., Schijven, D., van Rheenen, W., van Eijk, K.R., O'Brien, M., Kahn, R.S., Ophoff, R.A., Goris, A., Bradley, D.G., Al-Chalabi, A., et al. (2017). Genetic correlation between amyotrophic lateral sclerosis and schizophrenia. Nat Commun *8*, 14774.

Miki, H., Setou, M., Kaneshiro, K., and Hirokawa, N. (2001). All kinesin superfamily protein, KIF, genes in mouse and human. Proc. Natl. Acad. Sci. U.S.a. *98*, 7004–7011.

Millecamps, S., and Julien, J.-P. (2013). Axonal transport deficits and neurodegenerative diseases. Nat. Rev. Neurosci. *14*, 161–176.

Morton, S., Hesson, L., Peggie, M., and Cohen, P. (2008). Enhanced binding of TBK1 by an optineurin mutant that causes a familial form of primary open angle glaucoma. FEBS Lett. *582*, 997–1002.

Mucaki, E.J., Shirley, B.C., and Rogan, P.K. (2013). Prediction of Mutant mRNA Splice Isoforms by Information Theory-Based Exon Definition. Hum. Mutat. *270*, n/a–n/a.

Nakajima, K., Yin, X., Takei, Y., Seog, D.-H., Homma, N., and Hirokawa, N. (2012). Molecular Motor KIF5A Is Essential for GABA. Neuron *76*, 945–961.

Nishimura, A.L., Al-Chalabi, A., and Zatz, M. (2005). A common founder for amyotrophic lateral sclerosis type 8 (ALS8) in the Brazilian population. Hum Genet *118*, 499–500.

Nishimura, A.L., Mitne-Neto, M., Silva, H.C.A., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R.M., Gillingwater, T., Webb, J., et al. (2004). A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. The American Journal of Human Genetics *75*, 822–831.

Palomo, G.M., and Manfredi, G. (2015). Exploring new pathways of neurodegeneration in ALS: the role of mitochondria quality control. Brain Research *1607*, 36–46.

Patzkó, A., and Shy, M.E. (2011). Update on Charcot-Marie-Tooth disease. Curr Neurol Neurosci Rep *11*, 78–88.

Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet. Epidemiol. *32*, 381–385.

Reid, E., Kloos, M., Ashley-Koch, A., Hughes, L., Bevan, S., Svenson, I.K., Graham, F.L., Gaskell, P.C., Dearlove, A., Pericak-Vance, M.A., et al. (2002). A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10). The American Journal of Human Genetics *71*, 1189–1194.

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

Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., et al. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257–268.

Robberecht, W., and Eykens, C. (2015). The genetic basis of amyotrophic lateral sclerosis. Agg 327.

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., and Deng, H.X. (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature *362*, 59–62.

Salvi, E., Kutalik, Z., Glorioso, N., Benaglio, P., Frau, F., Kuznetsova, T., Arima, H., Hoggart, C., Tichet, J., Nikitin, Y.P., et al. (2012). Genomewide association study using a high-density single nucleotide polymorphism array and case-control design identifies a novel essential hypertension susceptibility locus in the promoter region of endothelial NO synthase. Hypertension *59*, 248–255.

Setou, M., Seog, D.-H., Tanaka, Y., Kanai, Y., Takei, Y., Kawagishi, M., and Hirokawa, N. (2002). Glutamate-receptor-interacting protein GRIP1 directly steers kinesin to dendrites. Nature *417*, 83–87.

Shatunov, A., Mok, K., Newhouse, S., Weale, M.E., Smith, B., Vance, C., Johnson, L., Veldink, J.H., van Es, M.A., van den Berg, L.H., et al. (2010). Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet Neurol *9*, 986–994.

Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R., and Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Research *34*, 3955–3967.

Smith, B.N., Ticozzi, N., Fallini, C., Gkazi, A.-S., Topp, S., Kenna, K.P., Scotter, E.L., Kost, J., Keagle, P., Miller, J.W., et al. (2014). Exome-wide Rare Variant Analysis Identifies TUBA4A Mutations Associated with Familial ALS. Neuron *84*, 324–331.

Smith, E.F., Shaw, P.J., and De Vos, K.J. (2017). ARTICLE IN PRESS. Neuroscience Letters 1–17.

Tanaka, Y., Kanai, Y., Okada, Y., Nonaka, S., Takeda, S., Harada, A., and Hirokawa, N. (1998). Targeted disruption of mouse conventional kinesin heavy chain, kif5B, results in abnormal perinuclear clustering of mitochondria. Cell *93*, 1147–1158.

Thiel, C., Kessler, K., Giessl, A., Dimmler, A., Shalev, S.A., Haar, von der, S., Zenker, M., Zahnleiter, D., Stöss, H., Beinder, E., et al. (2011). REPOR TNEK1 Mutations CauseShort-Rib Polydactyly Syndrome Type Majewski. The American Journal of Human Genetics *88*, 106–114.

van Es, M.A., Hardiman, O., Chio, A., Al-Chalabi, A., Pasterkamp, R.J., Veldink, J.H., and van den Berg, L.H. (2017). Amyotrophic lateral sclerosis. Lancet.

van Es, M.A., Veldink, J.H., Saris, C.G.J., Blauw, H.M., van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B., et al. (2009). Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat. Genet. *41*, 1083–1087.

van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L., Diekstra, F.P., Pulit, S.L., van der Spek, R.A.A., Võsa, U., de Jong, S., Robinson, M.R., et al. (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat. Genet. *48*, 1043–1048.

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., et al. (2009). Mutations in FUS, an RNA Processing Protein, Cause Familial Amyotrophic Lateral Sclerosis Type 6. Science *323*, 1208–1211.

Wang, L., and Brown, A. (2010). A hereditary spastic paraplegia mutation in kinesin-1A/KIF5A disrupts neurofilament transport. Mol Neurodegeneration *5*, 52.

Wang, W.-Y., Pan, L., Su, S.C., Quinn, E.J., Sasaki, M., Jimenez, J.C., Mackenzie, I.R.A., Huang, E.J., and Tsai, L.-H. (2013). Interaction of FUS and HDAC1 regulates DNA damage response and repair in neurons. Nat Neurosci *16*, 1383–1391.

Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., McKenna-Yasek, D., Baron, D.M., et al. (2012). Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature *488*, 499–503.

Xia, C.-H., Roberts, E.A., Her, L.-S., Liu, X., Williams, D.S., Cleveland, D.W., and Goldstein, L.S.B. (2003). Abnormal neurofilament transport caused by targeted disruption of neuronal kinesin heavy chain KIF5A. The Journal of Cell Biology *161*, 55–66.
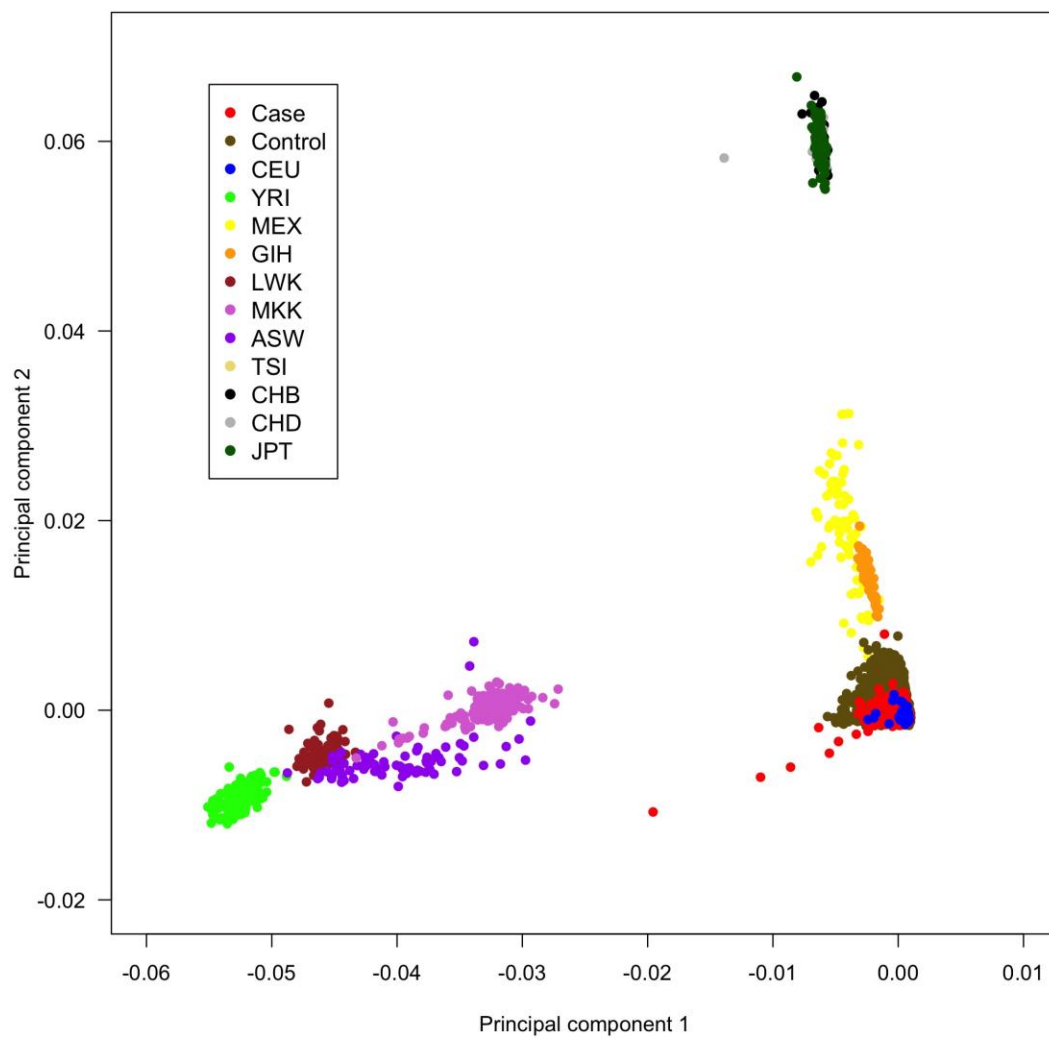
Figure S1. Multi-dimensional scaling plot of the 44,558 genotyped samples included in analysis compared to the HapMap populations.
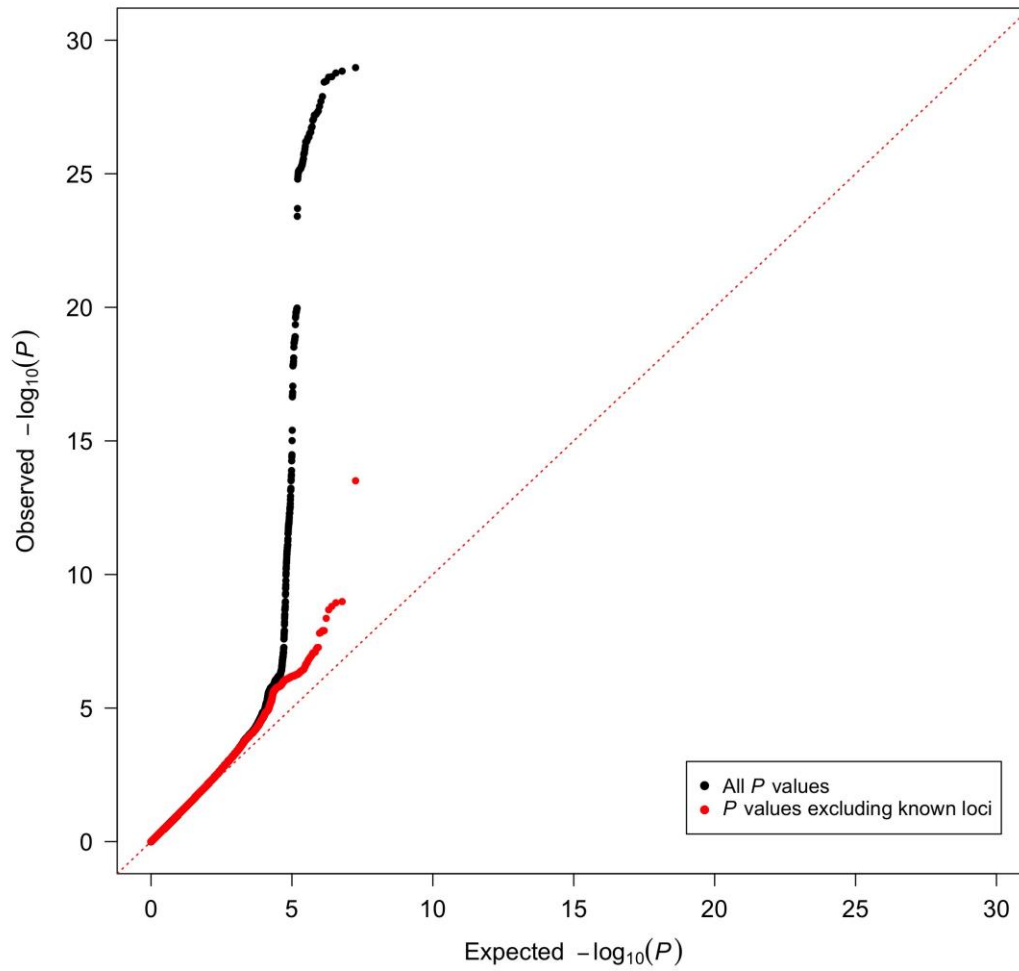
Figure S2. QQ plot of *P*-values from the meta-analysis based on logistic regression analysis. The black curve represents all SNPs, and the red curve represent SNPs after excluding variants within +/- 200 kilobases of C9orf72 and UNC13a loci. Raw lambda = 1.042 and adjusted lambda scaled to 1,000 cases and 1,000 controls = 1.001 based on the entire SNP dataset.
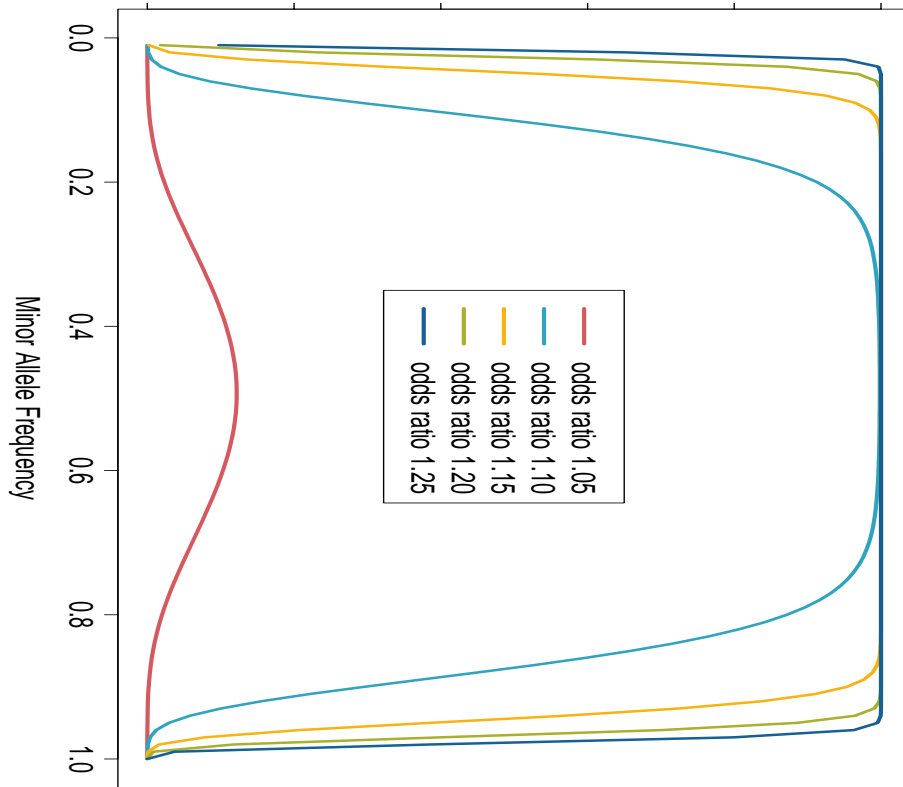
Figure S3. Power to detect associate loci for a cohort of 20,806 cases and 59,804 control subjects. Power calculations were performed using the QUANTO software program (version 1.2.4, available from: http://hydra.usc.edu/gxe) under an additive model assuming two-tailed P value of 5.0×10^{-8} (threshold P value for significance after Bonferroni correction for multiple testing), population risk of 0.0001 and complete linkage equilibrium between the genotyped SNP and the causative allele. Based on our findings for rs142321490 in the KIF5A locus on chromosome 12 (see **Table 1** in the main paper), a cohort of 20,806 cases and 59,804 controls has ~99.5% power to detect an associated SNP with a minor allele frequency of 0.018 and an odds ratio of 1.37 under the additive model.
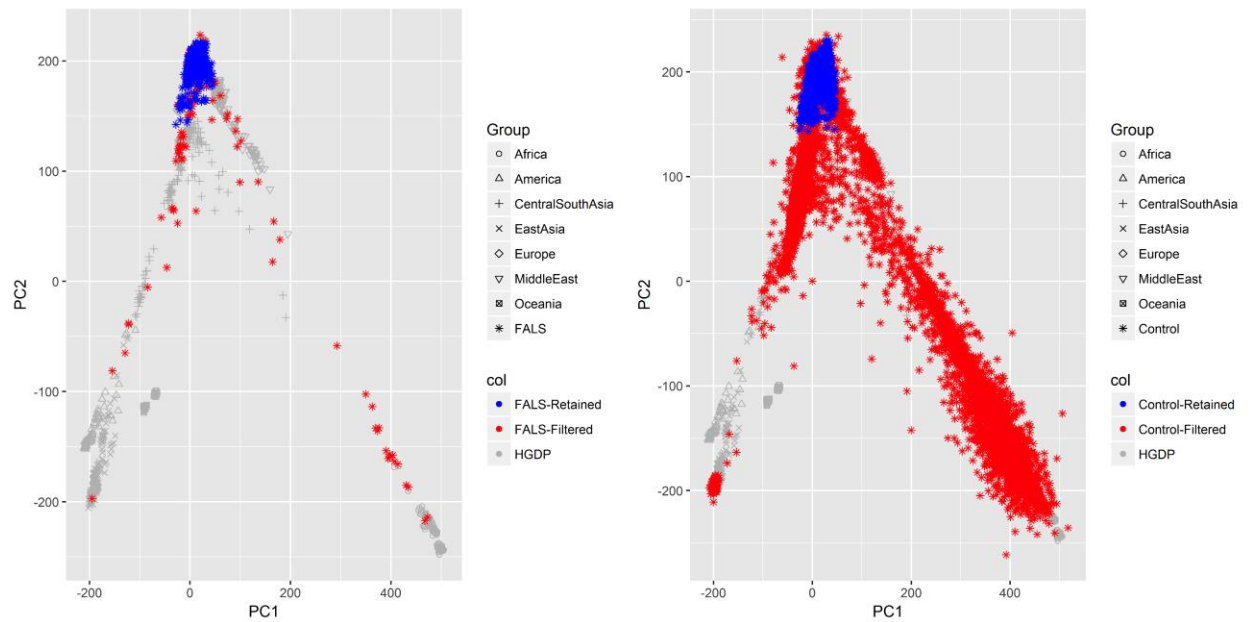
Figure S4. Principle components analysis of samples included in RVB analysis compared to the Human Diversity Panel. Ancestry filtering of FALS discovery cohort. LASER was used to generate PCA coordinates for samples from the Human genome diversity panel (HGDP). Samples from the FALS discovery cohort were then mapped to this reference co-ordinate space. The discovery cohort was restricted to cases and controls occurring within 3SD of mean for European HGDP samples along principle components 1-4.
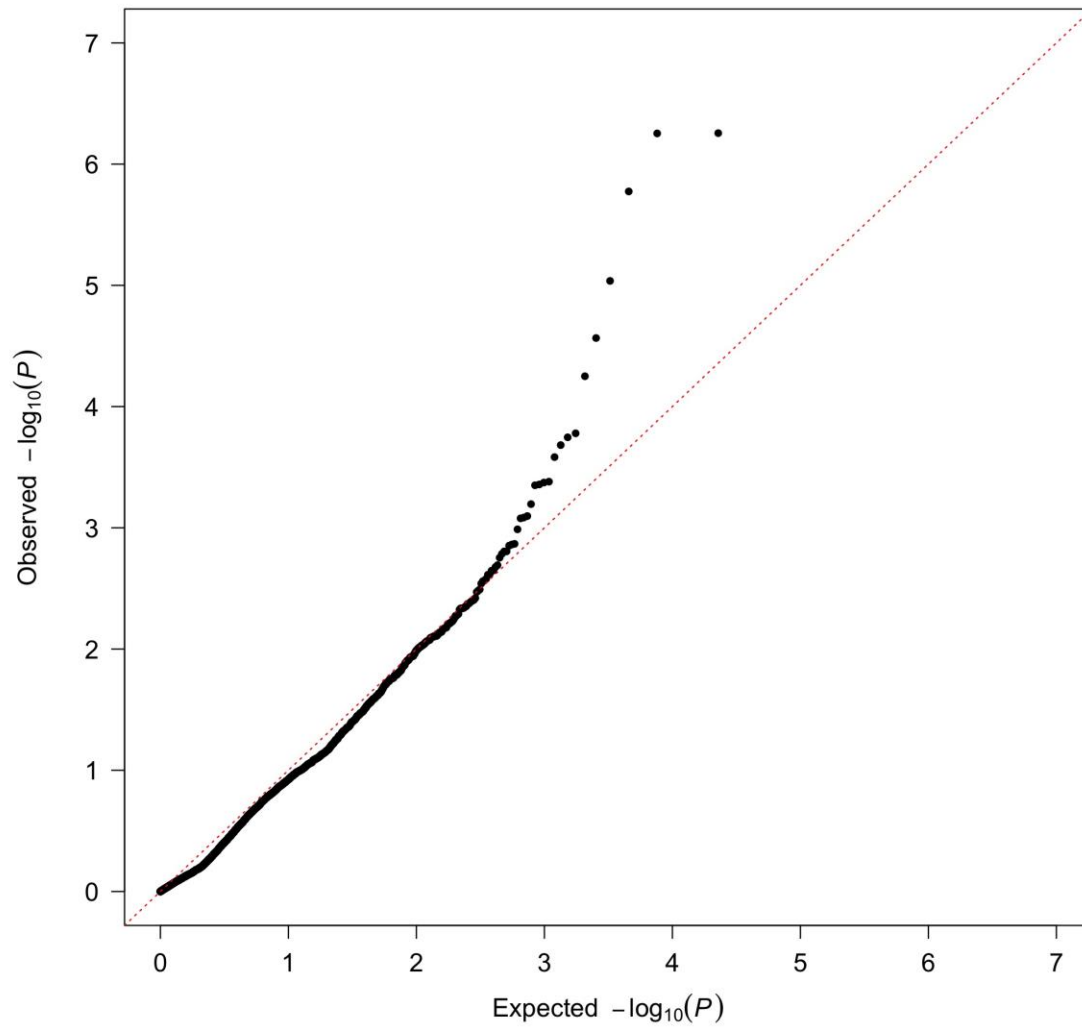
Figure S5. QQ plot of *P*-values from the gene-based rare variant burden analysis of exome data. The black curve represents all genes. The calculated genomic inflation factor ($\lambda = 0.93$) was calculated on the entire gene dataset.
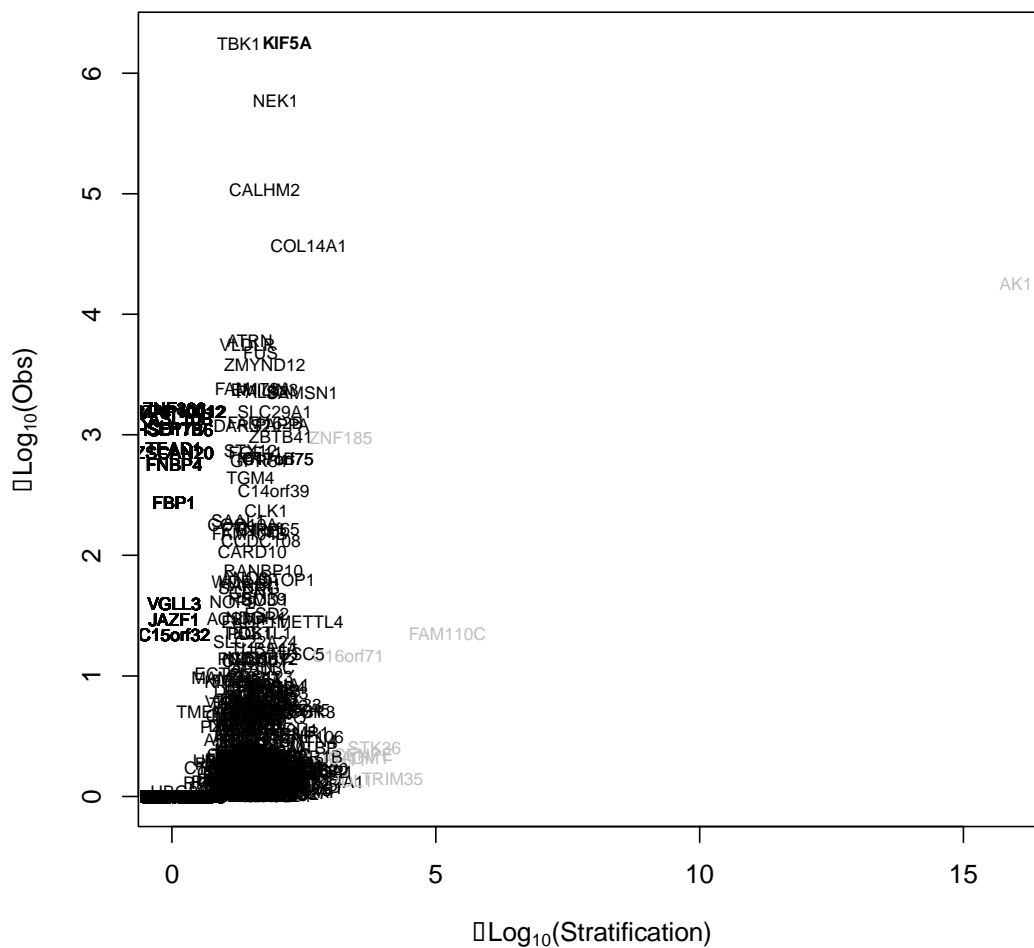
Figure S6. Control-control Analyses. P-values from RVB analysis of FALS cases vs controls (y-axis) are plotted against minimum p-values from RVB analyses of candidate batch effects (x-axis). To assess the potential impact of batch effects, the sample cohort was divided into 28 pseudo case-control groups 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.
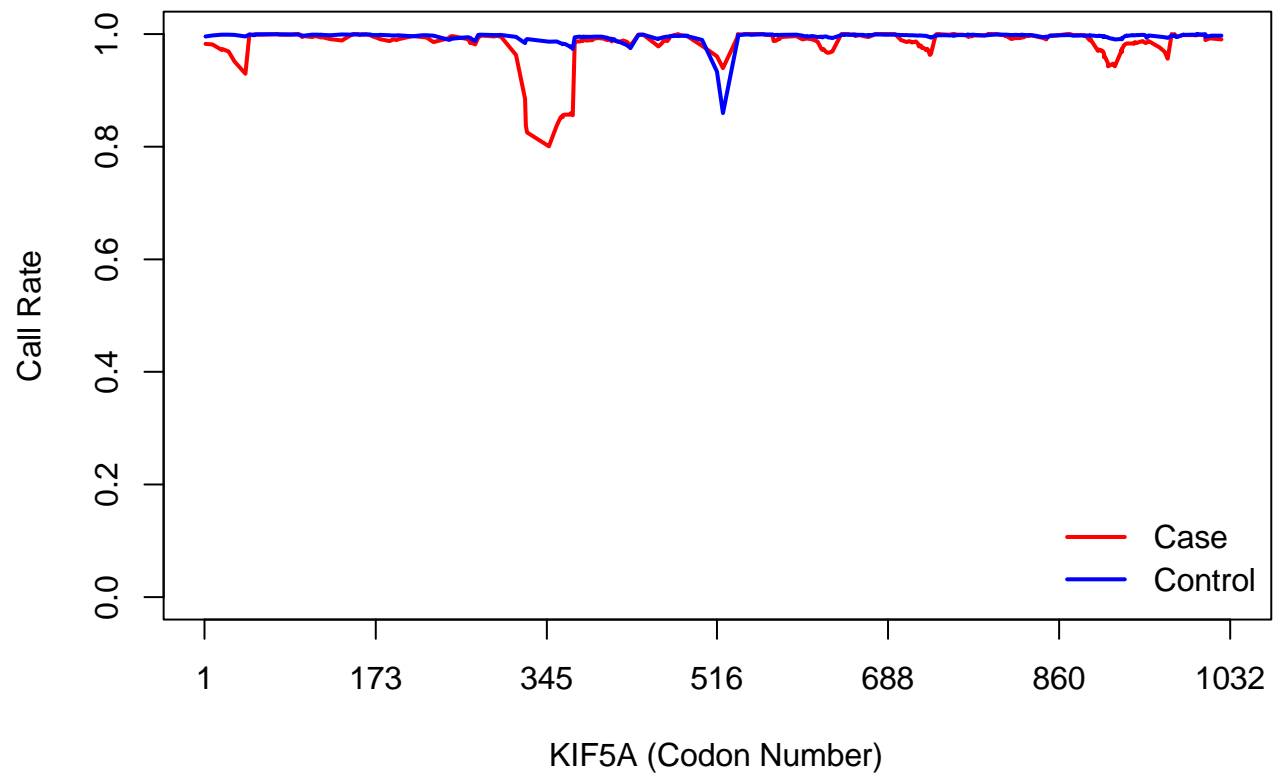
Figure S7. Plot of variant call rate across the KIF5A protein-coding region in FALS versus controls analyzed by RVB.
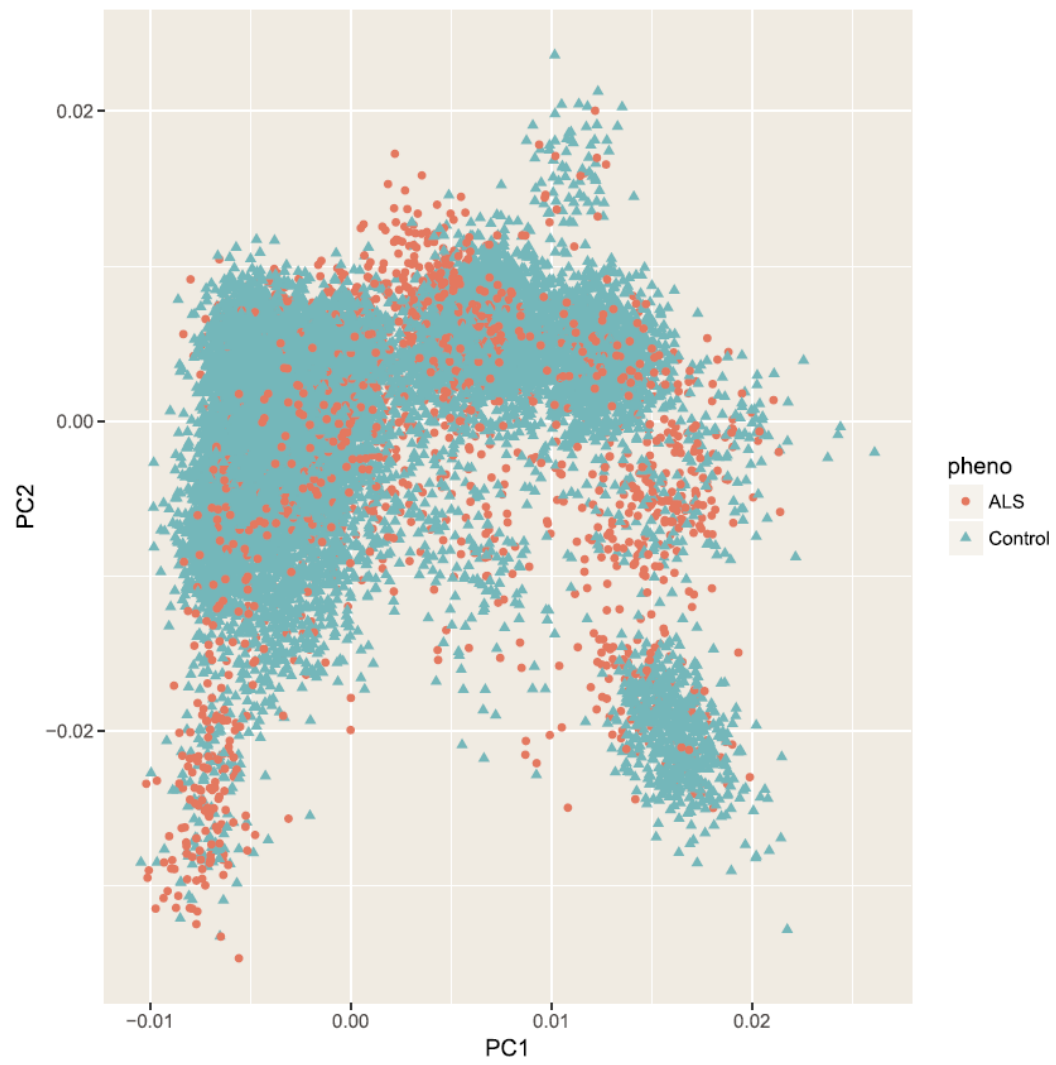
Figure S8. Principle Components analysis of samples included in *KIF5A* replication cohort.

Table S1. Demographics and baseline characteristics of patients diagnosed with ALS and control individuals included in analysis

	US		Italian		UK		French & Belgian		Total cohort	
	cases	controls	cases	controls	cases	controls	cases	controls	cases	controls
Sample number	3,777	33,365	2,853	2,143	449	226	1,150	595	8,229	36,329
Female (%)	1,515 (40.1)	23,870 (71.5)	1,239 (43.4)	896 (41.8)	193 (43.0)	109 (48.2)	486 (42.3)	422 (70.9)	3,433 (41.7)	25,297 (69.6)
Age (SD)	58.1 (12.3)	64.2 (13.3)	61.8 (11.8)	50.6 (17.4)	60.3 (12.8)	57.0 (0.0)	60.5 (12.6)	66.9 (16.8)	59.8 (12.3)	63.4 (13.9)
Bulbar-onset* (%)	963 (25.5)	-	741 (26.0)	-	141 (31.4)	-	357 (31)	-	2,202 (26.8)	-
Family history[†] (%)	458 (12.1)	-	248 (8.7)	-	54 (12.0)	-	195 (17.0)	-	955 (11.6)	-

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. DbGaP studies contributing to GWAS.

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

Table S3. SNPs achieving genome-wide significance based on logistic regression

	SNP Information			Present Study (8,229 Cases / 36,329 Controls)			Van Rheenen *et al.* (12,577 Cases / 23,475 Controls)			Combined Discovery Set (20,806 Cases / 59,804 Controls)			
SNP	Chr	Position	Gene	Beta [SE]	OR [95% CI]	*P*	Beta [SE]	OR [95% CI]	*P*	I^2	Beta [SE]	OR [95% CI]	*P*
Novel Loci													
rs117027576	12	57,316,603	*KIF5A*	0.373 [0.096]	1.45 [1.20-1.76]	1.1×10^{-4}	0.286 [0.070]	1.33 [1.16-1.53]	4.3×10^{-5}	25.6	0.316 [0.057]	1.37 [1.23-1.54]	2.3×10^{-8}
rs118082508	12	57,318,819	*KIF5A*	0.374 [0.096]	1.45 [1.20-1.76]	1.0×10^{-4}	0.288 [0.070]	1.33 [1.16-1.53]	3.8×10^{-5}	25.8	0.317 [0.051]	1.37 [1.23-1.54]	2.0×10^{-8}
rs113247976*	12	57,975,700	*KIF5A*	0.381 [0.086]	1.46 [1.23-1.74]	9.2×10^{-6}	0.288 [0.066]	1.33 [1.17-1.52]	1.1×10^{-5}	0.0	0.322 [0.052]	1.38 [1.24-1.53]	6.4×10^{-10}
rs116900480	12	58,656,105	*KIF5A*	0.354 [0.083]	1.42 [1.21-1.68]	1.9×10^{-5}	0.294 [0.065]	1.34 [1.18-1.53]	7.1×10^{-6}	0.0	0.317 [0.051]	1.37 [1.24-1.52]	6.6×10^{-10}
rs142321490	12	58,676,132	*KIF5A*	0.357 [0.082]	1.43 [1.21-1.68]	1.5×10^{-5}	0.292 [0.066]	1.34 [1.18-1.53]	8.0×10^{-6}	0.0	0.317 [0.056]	1.37 [1.24-1.52]	6.1×10^{-10}
Previously Published Loci													
rs10463311	5	150,410,835	*TNIP1*	-0.065 [0.024]	0.94 [0.89-0.98]	7.8×10^{-3}	-0.100 [0.020]	0.91 [0.87-0.94]	8.5×10^{-7}	0.0	-0.085 [0.016]	0.92 [0.89-0.95]	4.0×10^{-8}
rs3849943	9	27,543,382	*C9orf72*	-0.17 [0.024]	0.84 [0.80-0.88]	1.4×10^{-12}	-0.181 [0.020]	0.83 [0.80-0.87]	4.0×10^{-19}	0.0	-0.176 [0.016]	0.84 [0.81-0.86]	3.8×10^{-30}
rs74654358	12	64,881,967	*TBK1*	0.182 [0.058]	1.20 [1.07-1.34]	1.6×10^{-3}	0.206 [0.042]	1.23 [1.13-1.34]	7.7×10^{-7}	0.0	0.198 [0.034]	1.22 [1.14-1.30]	4.7×10^{-9}
rs12973192	19	17,753,239	*UNC13A*	-0.149 [0.026]	0.86 [0.82-0.91]	1.3×10^{-8}	-0.106 [0.019]	0.9 [0.87-0.93]	2.4×10^{-8}	38.6	-0.121 [0.015]	0.89 [0.86-0.91]	3.9×10^{-15}
rs75087725	21	45,753,117	*C21orf2*	0.687 [0.162]	1.99 [1.44-2.75]	2.2×10^{-5}	0.479 [0.074]	1.61 [1.39-1.87]	8.7×10^{-11}	31.1	0.515 [0.067]	1.67 [1.46-1.91]	1.8×10^{-14}

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. Suggestive SNPs with *P*-value less than 5.0x10^{-7} in genome-wide association analyses

SNP Information				Present Study (8,229 Cases / 36,329 Controls)				Van Rheenen *et al.* (12,577 Cases / 23,475 Controls)				Combined Discovery Set (20,806 Cases / 59,804 Controls)			
SNP	Chr	Position	Gene	Case MAF	Control MAF	OR [95% CI]	*P*	Case MAF	Control MAF	OR [95% CI]	*P*	Case MAF	Control MAF	OR [95% CI]	*P*
rs17070492	8	2,420,855	*LOC101927815*	10.01%	9.76%	1.03 [0.97-1.09]	3.4x10^{-1}	9.17%	10.09%	1.16 [1.09-1.23]	1.3x10^{-6}	9.50%	9.89%	1.13 [1.08-1.19]	1.0 x10^{-7}
rs10139154	14	31,147,498	*SCFD1*	34.10%	31.30%	1.13 [1.09-1.18]	4.1x10^{-12}	33.76%	31.17%	1.08 [1.04-1.12]	1.9x10^{-5}	33.90%	31.25%	1.08 [1.05-1.11]	1.4 x10^{-7}
rs10143310	14	92,540,381	*ATXN3*	24.85%	24.36%	1.03 [0.99-1.07]	1.9x10^{-1}	24.04%	22.95%	1.08 [1.04-1.13]	2.6x10^{-4}	24.36%	23.81%	1.09 [1.05-1.12]	3.2 x10^{-7}
rs9901522	17	14,673,934	*PMP22*	7.08%	6.31%	1.13 [1.06-1.21]	2.5x10^{-4}	6.87%	5.97%	1.16 [1.08-1.24]	4.6x10^{-5}	6.95%	6.18%	1.16 [1.10-1.22]	8.6 x10^{-8}

Table S5. Quality control filtering of the FALS Discovery and KIF5A replication Cohorts.

FALS discovery

Cohort	Cases	Controls
Initial Sample Set	1,463	41,410
Post HGDP Continental Ancestry Filter	1,397	24,563
Post Call Rate Filter	1,331	20,789
Post Heterozygosity Filter	1,319	20,664
Post Relatedness Filter	1,138	19,494

rs113247976 replication

Cohort	Cases	Controls
Initial Sample Set	12,415	21,533
Post Call Rate Filter	11,916	21,050
Post Heterozygosity Filter	11,721	21,028
Post Ancestry Filter (PCA)	11,373	21,009
Post Relatedness & GWAS Checksum Filter	4,160	18,650

LOF screen

Cohort	Cases	Controls
Initial Sample Set	12,415	21,533
Post Call Rate Filter	11,916	21,050
Post Heterozygosity Filter	11,721	21,028
Post Ancestry Filter (PCA)	11,373	21,009
Post Relatedness & FALS Discovery Overlap Filter	9,046	1,955

See Materials and Methods for details on filtering parameters.

Table S6. Clinical information of probands and relatives carrying *KIF5A* LOF Variants.

Position	Variant	Relation to Proband	DNA Available	Exon	cDNA	Description	Gender	Age of Onset (years)	Site of Onset	Survival (months)	Alive
57,975,729	GA>A	Proband	Y	26	c.2987delA	p.Asp996fs	M	45	n/a	n/a	n/a
57,976,382	C>T	Proband	Y	27	c.2993-3C>T	5' Splice Junction	M	29	L	>264	Y
57,976,382	C>T	Sister	Y	27	c.2993-3C>T	5' Splice Junction	F	52	L	84	N
57,976,382	C>T	Brother	Y	27	c.2993-3C>T	5' Splice Junction	M	18	L	324	N
		Brother	N				M	n/a	L	n/a	N
57,975,731	CA>C	Sporadic	Y	26	c.2989delA	p.Asn997fs	F	50	L	>96	Y
57,976,384	G>A	Sporadic	N	27	c.2993-1G>A	5' Splice Junction	n/a	52	B	n/a	n/a
57,976,385	GA>G	Proband	Y	27	c.2996delA	p.Asn999fs	M	42	L	>12	Y
		Brother	N				M	38	n/a	24	N
57,976,411	A>G	Proband	Y	27	c.3019A>G	p.Arg1007Gly	F	53	L	45	N
57,976,412	G>A	Proband	Y	27	c.3020G>A	p.Arg1007Lys	M	50	L	>108	Y
57,976,413	G>A	Proband	Y	27	c.3020+1G>A	3' Splice Junction	M	45	B	>220	Y
		Parent	N				n/a	47	n/a	156	N
		Uncle/Aunt	N				n/a	57	n/a	144	N
		Uncle/Aunt	N				n/a	55	n/a	121	N
		Uncle/Aunt	N				n/a	46	n/a	24	N
57,976,414	T>A	Proband	Y	27	c.3020+2T>A	3' Splice Junction	M	46	B	124	N
57,976,414	T>A	Brother	Y	27	c.3020+2T>A	3' Splice Junction	M	48	L	117	N
		Mother	N				F	35	L	144	N
57,976,415	A>G	Proband	Y	27	c.3020+3A>G	3' Splice Junction	M	50	B	54	N

All mutations were heterozygous; Genomic coordinates in NCBI build 37; Protein change based on transcript NM_004984.3;

Supplementary Table 7.

Analysis	FALS	Control	OR (95% CI)	P
Missense - Full CDS	9 (0.79%)	80 (0.41%)	1.93 (0.915-3.60)	8.09×10^{-2}
Missense - Motor Domain	3 (0.26%)	18 (0.09%)	3.27 (0.86-9.25)	7.74×10^{-2}
Missense - Microtubule Binding Domain	2 (0.18%)	8 (0.04%)	5.07 (0.95-18.52)	5.57×10^{-2}
Missense - Coiled-Coil Domain	3 (0.26%)	55 (0.28%)	1.01 (0.28-2.60)	9.83×10^{-1}
Missense - C-Terminal Domain	3 (0.26%)	7 (0.04%)	7.23 (1.74-24.55)	9.41×10^{-3}
Loss of Function	6 (0.53%)	3 (0.02%)	32.07 (9.05-135.27)	5.55×10^{-7}
Loss of Function (Including Frameshifts)	8 (0.70%)	3 (0.02%)	41.16 (12.61-167.57)	3.77×10^{-9}