Supplementary Materials:

Materials and Methods
FALS Sequencing Consortium Members
Acknowledgements
Figures S1-S4
Tables S1-S5

Other Supplementary Materials for this manuscript includes the following:
Table S6

Samples

Subjects for whole-exome analysis were drawn from IRB-approved genetic studies of ALS subjects at Consortium member institutions: the Columbia University Medical Center (which included Coriell samples), University of Massachusetts at Worchester, Stanford University (which included contributions from Emory University School of Medicine, the Johns Hopkins University School of Medicine, and the University of California, San Diego), Massachusetts General Hospital Neurogenetics DNA Diagnostic Lab Repository, Duke University, McGill University (which included contributions from Saint-Luc and Notre-Dame Hospital of the Centre Hospitalier de l'Université de Montréal (CHUM) (University of Montreal), Gui de Chauliac Hospital of the CHU de Montpellier (Montpellier University), Pitié Salpêtrière Hospital, Fleurimont Hospital of the Centre Hospitalier Universitaire de Sherbrooke (CHUS) (University of Sherbrooke), Enfant-Jésus Hospital of the Centre hospitalier affilié universitaire de Québec (CHA) (Laval University), and Montreal General Hospital and Montreal Neurological Institute and Hospital of the McGill University Health Centre), and Washington University in St. Louis (which included contributions from Houston Methodist Hospital, Virginia Mason Medical Center, University of Utah, and Cedars Sinai Medical Center). Subjects for follow-up sequencing came from the same centers plus the University of Pennsylvania and University of Amsterdam. Genotypes for the 51 genes used in the replication analysis were also provided for FALS exomes sequenced as previously described(*13*). All patients were

diagnosed according to El Escorial revised criteria as suspected, possible, probable, or definite ALS by neuromuscular physicians. Subjects were considered sporadic if no first or secondary relatives had been diagnosed with ALS or died of an ALS-like syndrome. Details are presented in Table S1.

All samples known to be carriers of the *C9orf72* expansion were excluded from all analyses. There were 886 case exomes that were not screened for this variant. Additionally, prior to exome sequencing, some samples were screened for variants in known ALS genes and were only sequenced if they were found to be negative for a mutation in that gene. The number of pre-screened discovery samples for each gene were 430 for *SOD1*, 334 for *TARDBP* and *FUS*, and 143 for *VAPB, DCTN1, ANG, FIG4, OPTN, VCP, UBQLN2, EWSR1, DAO, SQSTM1, SETX*, and *TAF15*. The 543 exomes used in the replication stage were also screened for variants in *TARDBP, FUS, SOD1, VCP, PFN1*, and *TUBA4A* prior to use in this study.

Control samples were sequenced as part of other studies at the Duke CHGV, HudsonAlpha, and McGill University and were not enriched for (but not specifically screened for) ALS or other neurodegenerative disorders. Control samples were matched to case samples in terms of similar capture kits and coverage levels (Tables S2 and S3). Except for the exome cases used in the replication phase, all samples used within each analysis subset were processed using identical pipelines.

Only genetically European ethnicity samples were included in the analysis. Samples were also screened with KING(*46*) to remove duplicate samples between the custom capture and exome datasets and to remove second-degree or higher relatives in the exome datasets; exomes with incorrect sexes according to X:Y coverage ratios were removed, as were contaminated samples according to VerifyBamID(*47*).

Sequencing

Sequencing of DNA was performed at the Duke Center for Human Genome Variation, McGill University, Stanford University, HudsonAlpha, and University of Massachusetts,

Worcester. Samples were either exome sequenced using the Agilent All Exon (37MB, 50MB or 65MB) or the Nimblegen SeqCap EZ V2.0 or 3.0 Exome Enrichment kit or whole-genome sequenced using Illumina GAIIx or HiSeq 2000 or 2500 sequencers according to standard protocols (see Table S2). Follow-up custom capture sequencing was performed using the same methods as the exome sequencing with the Nimblegen SeqCap EZ Choice to an average coverage of 144.60x within the capture regions, with an average of 99.37% bases covered at least 5x.

Case and control samples were processed at the Duke CHGV (discovery Duke and McGill/Stanford datasets and replication custom capture dataset), HudsonAlpha (discovery HudsonAlpha dataset) and University of Massachusetts, Worcester (replication exome dataset) as follows. The Illumina lane-level fastq files were aligned to the Human Reference Genome (NCBI Build 37) using the Burrows-Wheeler Alignment Tool (BWA). We then used Picard software to remove duplicate reads and process these lane-level SAM files, resulting in a sample-level BAM file that is used for variant calling. We used GATK to recalibrate base quality scores, realign around indels, and call variants. The Duke, McGill/Stanford, custom capture and replication exome variants were required to have a quality score (QUAL) of at least 20 (30 for replication exomes), a genotype quality (GQ) score of at least 20, and at least 10x coverage. Additionally, Duke, McGill/Stanford and custom capture variants were required to have a quality by depth (QD) score of at least 2 and a mapping quality (MQ) score of at least 40. For Duke and McGill/Stanford exomes and custom capture samples, indels were also required to have a maximum strand bias (FS) of 200 and a minimum read position rank sum (RPRS) of -20. Other variants were restricted according to VQSR tranche (calculated using the known SNV sites from HapMap v3.3, dbSNP, and the Omni chip array from the 1000 Genomes Project): the cutoffs for Duke, McGill/Stanford and custom capture variants were a tranche of 99.9% for SNVs in genomes and exomes and 99% for indels in genomes; the cutoffs for HudsonAlpha were a 99% tranche for SNVs and 95% tranche for indels; and the cutoff for the replication exomes was a 97% tranche for SNVs and indels. Variants were excluded if they were determined to be sequencing, batch-specific or kit-specific

artifacts or if they were marked by EVS as being failures in the Duke, McGill/Stanford and custom capture datasets. Variants were annotated to Ensembl 73 using SnpEff.

Predisposition analysis

This study first analyzed whole-exome sequence data for discovery purposes and then performed follow-up custom capture sequencing of 51 genes and interrogated these 51 genes in additional case exomes. We analyzed the discovery samples in separate groups to control for differences in sequencing methods and coverage levels. The Duke analysis used genomes and Nimblegen and Agilent 65MB exomes with at least 90% of the consensus coding sequence (CCDS) bases covered to at least 10x, the HudsonAlpha analysis used Nimbelgen exomes with at least 90% of the CCDS bases covered to at least 10x, and the McGill/Stanford University analysis used genomes and Agilent 37MB and 50MB exomes with 75% of the CCDS covered at least 10x.

Our study focused on gene-based collapsing analyses. First, the number of bases with at least 10x coverage was calculated for each CCDS exon plus 10 bp into each intron for each sample. Because difference in coverage can cause biased results, exons with coverage differences (cutoff tailored to each analysis (see Table S3)) between cases and controls were excluded from analysis.

For each gene, each sample was then indicated as carrying or not carrying a qualifying variant. Qualifying variants were defined for dominant (one qualifying variant per gene; minor allele frequency (MAF) cutoff of 0.05% internally and 0.005% in ExAC) and recessive (two qualifying variant per gene, including homozygous and potentially compound heterozygous samples as carriers; MAF cutoff 1%) models. These allele frequency thresholds used a leave-one-out method for the combined sample of cases and controls, in each analysis group (where the MAF of each variant was calculated using all samples except for the sample in question). Variants were also required to pass this MAF cutoff in the publically available ExAC global frequencies; HudsonAlpha analyses additionally required a MAF below 0.01 in the 1000 Genomes Data (*41, 48*). We performed analyses of CCDS genes using three methods to identify qualifying variants:

1) all non-synonymous and canonical splice variants (coding model), 2) all non-synonymous coding variants except those predicted by PolyPhen-2 HumVar(*49*) to be benign (no benign model), and 3) only stop gain, frameshift and canonical splice variants (loss-of-function [LoF] model). Qualifying variants were identified using Analysis Tools for Annotated Variants (http://humangenome.duke.edu/software) at Duke and an in-house pipeline at HudsonAlpha.

The total number of cases and controls with qualifying variants in each gene for each model were calculated, and a Cochran-Mantel-Haenszel (CMH) test was performed in R to generate a combined, stratified p-value across all three discovery groups. Genes were only considered if they were assessed in all three discovery cohorts and had more than one case or control sample with a genetic variant meeting the inclusion criteria for at least one of the genetic models tested. A Breslow-Day test was applied to assess homogeneity of effects across different groups, and $p>0.05$ was required for the gene to be considered. The adjusted alpha after correcting for the number of genes tested over all ?six? genetic models becomes $p<9.07\times10^{-7}$. The p-values in the replication dataset were calculated separately, and combined p-values using the discovery and replication dataset were also calculated using a Cochran-Mantel-Haenszel test.

The 51 genes for custom capture sequencing follow-up were chosen by identifying genes that were of statistical and/or biological interest in initial exome sequencing results utilizing 2,843 cases and 4,310 controls and employing both a singleton and 0.5% MAF dominant model as well as a 1% MAF recessive model (Table S4). Specifically, after restricting to genes below the 50^{th} percentile for intolerance to functional mutation(*50*), removing associations driven by artifacts, and requiring signals to be consistent between the HudsonAlpha and Duke analyses and not supporting control enrichment in the McGill/Stanford University analysis, we chose the 19 genes with the lowest p-values by any model, the 23 genes with the lowest p-values according to the singleton not benign model, the 4 genes with the lowest p-values in any recessive model, the 8 genes with the lowest p-values in any LoF model, and also *KCNT1*, *UBE2D2*, and *SCEL*, which

received encouraging association statistics but did not meet the above criteria. Overlap between the lists resulted in a total of 51 genes. The processing of custom capture samples was performed as described above for the Duke datasets.

Clinical phenotype analysis

Detailed phenotypic information was available from the medical records of a subset of the participants. We focused on the clinical phenotypes of age at onset (n=2,190 cases), site of onset (n=2,096 cases; bulbar vs. spinal), and survival time in months from age at onset (n=1,828 cases).

Carriers of qualifying variants in each gene were identified using the same method as that described above, but here the dominant model used a 0.5% instead of 0.05% MAF cutoff, and MAF screening used the information from the European and African EVS samples (*51*). Linear regression analysis was performed to analyze age at onset, Firth logistic regression was used to analyze the site of onset, and Cox Proportional-Hazards Regression was used to analyze survival in R(*52*). These analyses always included sex, analysis group and EIGENSTRAT axes (which were calculated for EIGENSTRAT-pruned whites only and were created using the genotypes for variants from the Illumina HumanCore chip that overlap exons and were not found to be influenced by sequencing or genotyping method) as covariates, and the analysis of survival additionally included age at onset as a covariate and required at least 1% of cases to have variants in a given gene for it to be included in the analysis (with the exception of previously reported ALS genes, which we analyzed separately and included regardless of the number of carriers).

Cell Culture, Transfection, and Reagents

All cell lines were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FBS) and maintained at 5% CO_2/37^{O}C. Plasmid were transfected using lipid-based reagents (Lipofectamine 2000). Lentiviruses were made in 293T cells and used to infect 293T cells followed by selection on puromycin.

Immunoprecipitation and Proteomic Analysis

AP-MS and *CompPASS* analysis using the (Comparative Proteomics Analysis Software Suite) were performed as previously described(*53, 54*). Briefly, 10^7 cells were lysed in lysis buffer [50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% NP-40 and supplemented with protease inhibitors (Roche)] for 30 min on ice to obtain whole cell extracts. Lysates were incubated with 30 ml of anti-HA resin (Covance) and after extensive washing with lysis buffer, proteins were eluted with HA peptide prior to trichloroacetic acid precipitation, trypsinization, and stage tipping. Samples were ran in technical duplicate on an LTQ Velos (Thermo) mass spectrometer, and spectra search with Sequest prior to target-decoy peptide filtering, and linear discriminant analysis(*55*). Protein Assembler was used to convert spectral counts to average protein spectral matches (APSMs), which takes into account peptides, which match more than one protein in the database. Peptides were identified with a false discovery rate of < 1.0% and the protein false discovery rate was <4.85% (Table S5). The following MS^2 conditions were used: Activation Type – Collision induced dissociation; Minimum Signal Required - 2000.0; Isolation width (m/z) - 1.00; Normalized Collision Energy - 35.0; Default Charge State – 2; Activation Q - 0.250; Activation Time (ms) - 10.000. Peptide data (APSMs) were uploaded into the *CompPASS* algorithm housed within the CORE environment. For *CompPASS* analysis, we employed a stats table of 170 unrelated bait proteins analyzed in an analogous manner, including deubiquitinating enzymes and autophagy components(*53, 54*). The *CompPASS* system identifies high confidence candidate interacting proteins (HCIPs) based on the normalized weighted D (NWD)-score, which incorporates the frequency with which they identified within the stats table, the abundance (APSMs, average peptide spectral matches) when found, and the reproducibility of identification in technical replicates, and also determines a z-score based on APSMs(*53, 54*). Proteins with NWD-scores >1.0 are considered HCIPs, although we also note that some proteins that may be *bona fide* interacting proteins may not reach the strict threshold set by a NWD-score of > 1.0.

For IP of endogenous NEK1, VAPB or ALS2 in NSC-34 cells expressing either HA-ALS2, HA-VAPB or FLAG-NEK1, respectively, ~0.5 mg of lysate was incubated with 0.5 µg of the indicated antibody (anti-HA Abcam ab9110, anti-Nek1 Bethyl Labs A304-570A, anti-FLAG Sigma F1804, anti-ALS2 Sigma SAB4200137, anti-VAPB Bethyl Labs A302-894A) or control IgG (Cell Signaling Technologies) overnight at 4°C. Protein G resin (25 µl) was then added to the IP reaction and incubated for a further 2 hours at 4°C. Beads were washed three times with lysis buffer. After washing, 4x SDS loading buffer was added and the samples were boiled for 5 min. Samples were separated on a SDS-PAGE gel prior to immunoblot analysis according to standard procedures using primary antibodies at 1:1000 overnight at 4°C, HRP-conjugated secondary antibodies (Promega) at 1:5000 for 1 hour at room temperature, and chemiluminescent detection (PerkinElmer).

FALS Sequencing Consortium

Other FALS Sequencing Consortium Members are as follows:

Orla Hardiman[1], Russell L McLaughlin[1], Letizia Mazzini[2], Ian P Blair[3], Kelly L Williams3, Garth A Nicholson4, Safa Al-Sarraj5, Andrew King5, Emma L Scotter5, Simon Topp5, Claire Troakes5, Caroline Vance5, Sandra D'Alfonso6, Stefano Duga7, Lucia Corrado8, Anneloor LMA ten Asbroek9, Daniela Calini10, Claudia Colombrita11, Antonia Ratti11, Cinzia Tiloca12, Vincenzo Silani13, Seneshaw Asress14, Meraida Polak14, Frank Diekstra15, Wouter van Rheenen15, Eric W Danielson16, Claudia Fallini16, Pamela Keagle16, Elizabeth A Lewis16, Jason Kost17, Gianni Sorarù18, Cinzia Bertolin19, Giorgia Querin19, Cristina Cereda20, Stella Gagliardi20, Mauro Ceroni21, Giuseppe Lauria22, Jacqueline de Belleroche23, Giacomo P Comi24, Stefania Corti24, Roberto Del Bo24, Martin R Turner25, Kevin Talbot26, Hardev Pall27, Karen E Morrison28, Pamela J Shaw29, Jesús Esteban-Pérez30, Alberto García-Redondo30, José Luis Muñoz-Blanco31, Barbara Castellotti32, Cinzia Gellera32, Viviana Pensato32, Franco Taroni32, Zheyang Wu33.

[1]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Republic of Ireland.

[2]ALS Center, Department of Neurology, 'A. Avogadro' University of Eastern Piedmont, Novara, Italy

[3]Australian School of Advanced Medicine, Macquarie University, Sydney, NSW 2109, Australia

[4]Australian School of Advanced Medicine, Macquarie University, Sydney, NSW 2109, Australia; Northcott Neuroscience Laboratory, ANZAC Research Institute, Sydney, NSW 2139, Australia

[5]Centre for Neurodegeneration Research, King's College London, Department of Clinical Neuroscience, Institute of Psychiatry, Psychology & Neuroscience, London, SE5 8AF, UK

[6]Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases (IRCAD), ''A. Avogadro'' University, 28100 Novara, Italy

[7]Department of Medical Biotechnology and Translational Medicine - Università degli Studi di Milano, Milan, Italy.

[8]Department of Medical Sciences, 'A. Avogadro' University of Eastern Piedmont, Novara, Italy

[9]Department of Neurogenetics and Neurology, Academic Medical Centre, Amsterdam, The Netherlands

[10]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, 20149 Milan, Italy

[11]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, 20149 Milan, Italy; Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center - Università degli Studi di Milano, Milan 20122 Italy.

[12]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, 20149 Milan, Italy; Doctoral School in Molecular Medicine, Department of Sciences and Biomedical Technologies, Universita` degli Studi di Milano, Milan 20122, Italy.

[13]Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milan 20149, Italy.; Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center - Università degli Studi di Milano, Milan 20122 Italy.

[14]Department of Neurology, Emory University, Atlanta, GA 30322, USA.

[15]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, 3508 GA Utrecht, The Netherlands

[16]Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.

[17]Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA.; Worcester Polytechnic Institute, Worcester, MA 01609, USA.

[18]Department of Neurosciences, University of Padova, Padova, Italy

[19]Department of Neurosciences, University of Padova, Padova, Italy.

[20]Experimental Neurobiology Laboratory, IRCCS 'C. Mondino' National Neurological Institute, 27100 Pavia, Italy

[21]Experimental Neurobiology Laboratory, IRCCS 'C. Mondino' National Neurological Institute, 27100 Pavia, Italy; Department of Neurological Sciences, University of Pavia, 27100 Pavia, Italy

[22]Headache and Neuroalgology Unit, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan 20133, Italy.

[23]Neurogenetics Group, Division of Brain Sciences, Hammersmith Hospital Campus, Burlington Danes Building, Du Cane Road, London W12 0NN

[24]Neurology Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan 20122, Italy; Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center - Università degli Studi di Milano, Milan 20122 Italy.

[25]Nuffield Department of Clinical Neurosciences, University of Oxford, UK

[26]Nuffield Department of Clinical Neurosciences, University of Oxford, UK; Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Oxford, UK

[27]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, The University of Birmingham, Birmingham, UK

[28]School of Clinical and Experimental Medicine, College of Medical and Dental Sciences, University of Birmingham, UK.; Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust UK.

[29]Sheffield Institute for Translational Neuroscience, University of Sheffield, Sheffield, UK

[30]Unidad de ELA, Instituto de Investigación Hospital 12 de Octubre de Madrid, SERMAS, and Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER U-723), Madrid, Spain.

[31]Unidad de ELA, Instituto de Investigación Hospital Gregorio Marañón de Madrid, SERMAS, Spain.

[32]Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milan 20133, Italy.

[33]Worcester Polytechnic Institute, Worcester, MA 01609, USA.

Acknowledgements:

Funding for this study was provided by Biogen Idec MA, Inc. P.C.S. was supported through the auspices of Dr. H. Robert Horvitz, an Investigator at the Howard Hughes Medical Institute in the Department of Biology at the Massachusetts Institute of Technology. J.W.H. was supported by a research grant from Biogen-Idec and R37 NS083524. The Shaw laboratory is supported by grants from the Wellcome Trust and Medical Research Council Grants 089701/Z/09/Z, G0900688 and MR/L021803/1, MND Association and American ALS Association. Funding for the FALS Sequencing Consortium was provided by the National Institutes of Health (NIH)/National Institute of

Neurological Disorders and Stroke (NINDS) (5RO1NS050557; 5R01 NS067206; 5R01NS065847; 1R01NS073873; 5R01NS079836 R01NS065847; and R01NS073873 [J.E.L., R.H.B.]), the American ALS Association (N.T., V.S., C.E.S., J.E.L., R.H.B.), Project MinE, the MND Association (N.T., V.S., A.A.C, C.E.S., J.E.L.), the Angel Fund (R.H.B.), Project ALS/P2ALS (R.H.B.), the ALS Therapy Alliance (R.H.B., J.E.L.), the Pierre L. de Bourghknecht ALS Research Foundation (R.H.B.), a Francesco Caleffi donation (N.T., V.S.), the Medical Research Council, the Heaton-Ellis Trust and AriSLA cofinanced with support of ‘‘5 x 1000’’—Healthcare research of the Italian Ministry of Health (grants EXOMEFALS 2009 and NOVALS 2012 [N.T., C.G., C. Tiloca, V.S., J.E.L.]), the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit at South London (C.E.S., A.A.C.), Maudsley NHS Foundation Trust (C.E.S., A.A.C.), King’s College London (C.E.S., A.C.), the Motor Neurone Disease Research Institute of Australia (Leadership Grant to IPB and a Bill Gole fellowship to K.L.W.), United Kingdom Medical Research Council under the aegis of JPND – http://www.jpnd.eu (A.A.C.), the National Health and Medical Research Council of Australia (1004670), Canadian Institutes of Health Research (208973, G.A.R.), Muscle Dystrophy Association (153959, G.A.R.), Instituto de Salud Carlos III - ISCIII (grants EC08/00049; PI10/00092; PI12/ 03110 – Erare european call – J.E.P., A.G.R.), FUNDELA (Spanish foundation for the development of ALS research, J.E.P, A.G.R.) and the Mireia Barneda project ‘‘No llores, no te rindas’’ (J.E.P., A.G.R.), the Netherlands ALS Foundation. Research leading to these results has received funding from the European Community's Health Seventh Framework Programme (FP7/2007-2013).

We acknowledge Jarod Shelton at Washington University St. Louis, Peggy Allred at Cedars Sinai, Luis Lay and Sharon Halton at Houston Medical, Karla Figueroa at University of Utah, Michael Baughn at University of California, San Diego, and Marlyne Silver, Karen Grace, and Ken Cronin at Duke University for recruiting subjects, preparing DNA, and maintaining clinical databases. We acknowledge M.T. van Meegen and A.P. Prijs at University of Amsterdam and Mei Liu and Paul Carmillo at Biogen Idec for technical assistance.

DNA panels from the NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds) were used in this study, as well as clinical data. The submitters that contributed samples are acknowledged in detailed descriptions of each ALS1 sample.

We would like to acknowledge the following individuals for the contributions of control samples: Dr. Joseph McEvoy, Dr. Anna Need, Mr. Jordan Silver and Ms. Marlyne Silver; Dr. Deborah Koltai Attix, and Ms. Jill McEvoy, Dr. Kenneth Schmader, Dr. Shelley McDonald, Dr. Heidi K. White, Dr. Mamata Yanamadala, and the Carol Woods and Crosdaile Retirement Communities; Dr. Gianpiero Cavalleri, Dr. Norman Delanty, and Dr. Chantal Depondt; Dr. William B. Gallentine, Dr. Erin L. Heinzen , Dr. Aatif M. Husain, Ms. Kristen N. Linney, Dr. Mohamad A. Mikati, Dr. Rodney A. Radtke, Dr. Saurabh R. Sinha, and Ms. Nicole M. Walley; Dr. Ruth Ottman; Dr. Julie Hoover-Fong, Dr. Nara L. Sobreira and Dr. David Valle; Dr. Yong-Hui Jiang; Dr. Demetre Daskalakis; Dr. William L. Lowe; Dr. James Burke, Dr. Christine Hulette, and Dr. Kathleen Welsh-Bohmer; Dr. Vandana Shashi and Ms. Kelly Schoch; Mr. David H. Murdock and The MURDOCK Study Community Registry and Biorepository; Dr. Scott M. Palmer; Dr. Zvi Farfel, Dr. Doron Lancet, and Dr. Elon Pras; Mr. Arthur Holden and Dr. Elijah Behr; Dr. Annapurna Poduri; Dr. M Chiara Manzini; Dr. Nicole Calakos; Dr. Patricia Lugar; Dr. Doug Marchuk; Dr. Qian Zhao; Dr. Sarah Kerns and Dr. Harriet Oster; Dr. Rasheed Gbadegesin and Dr. Michelle Winn; Dr. Francis J. McMahon and Nirmala Akula; Dr. Eli J. Holtzman; Dr. Joshua Milner; Dr. Deborah Levy; Dr. Ann Pulver; and Dr. Michael Hauser.

The collection of control samples was funded in part by Bryan ADRC NIA P30 AG028377, NIMH Award RC2MH089915, NINDS Award RC2NS070344, NIAID Award R56AI098588, the Ellison Medical Foundation New Scholar award AG-NS-0441-08, an award from SAIC-Frederick, Inc. (M11-074), NIMH Award K01MH098126, MH057314 and MH068406 to Ann E. Pulver, Sc.D. (Johns Hopkins University School of Medicine), and the Epi4K Gene Discovery in Epilepsy study (NINDS U01-NS077303) and the Epilepsy Genome/Phenome Project (EPGP - NINDS U01-NS053998). The collection of control samples was supported

in part by funding from the Division of Intramural Research, NIAID, NIH, and with federal funds by the Center for HIV/AIDS Vaccine Immunology ("CHAVI") under a grant from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Grant Number UO1AIO67854.

J.W.H. is a consultant for Biogen-Idec and Millennium: the Takeda Oncology Company. R.B. is a consultant for Biogen, Idec and a cofounder of AviTx. F.B. is a founder of Regenesance.

The results presented in this study can be found in the Supplementary Materials.

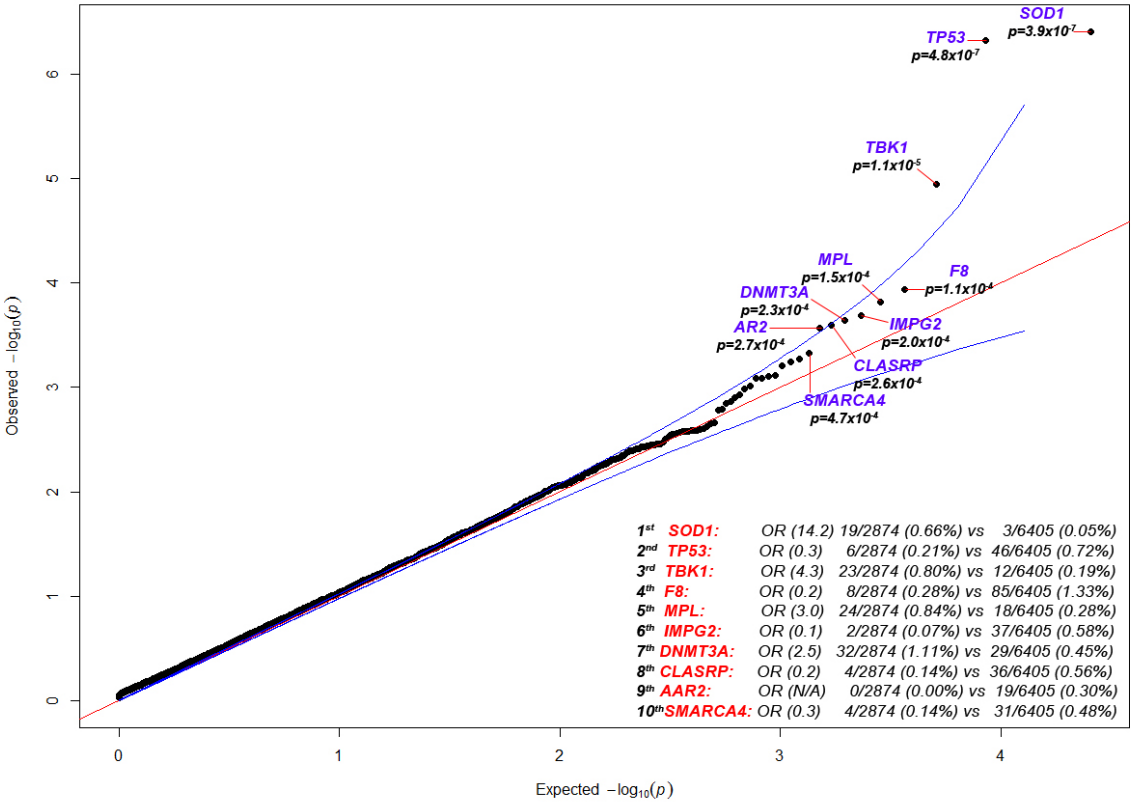

Fig. S1. QQ plot of discovery results for dominant no benign model

Shown are the results for the analysis of 2,874 case and 6,405 control exomes. There were 16,335 covered genes passing QC with more than one case or control carrier for this test, and the genes with the top 10 associations are labeled. The lambda quantifying inflation is 1.060. The association with *SOD1* passes correction for multiple tests.

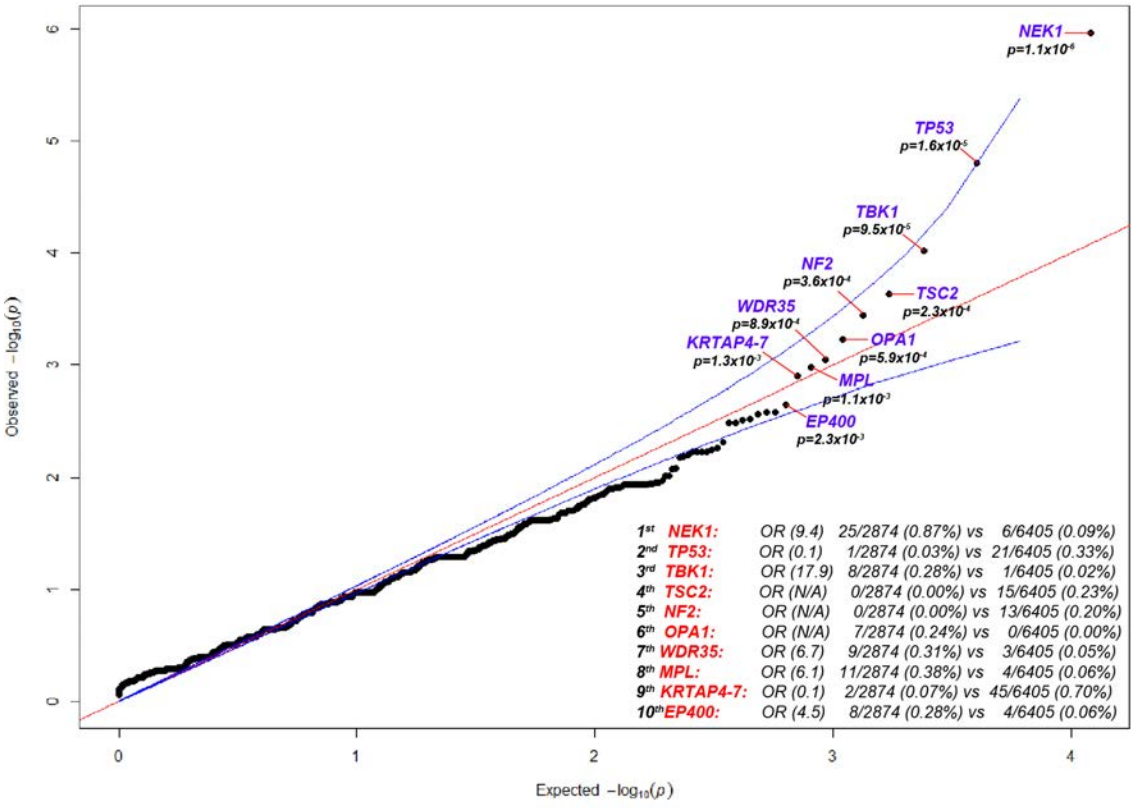

Fig. S2 QQ plot for the discovery results from the dominant LoF model

Shown are the results for the analysis of 2,874 case and 6,405 control exomes. There were 9,816 covered genes passing QC with more than one case or control carrier for this test, and the genes with the top 10 associations are labeled. The lambda quantifying inflation is 0.958.

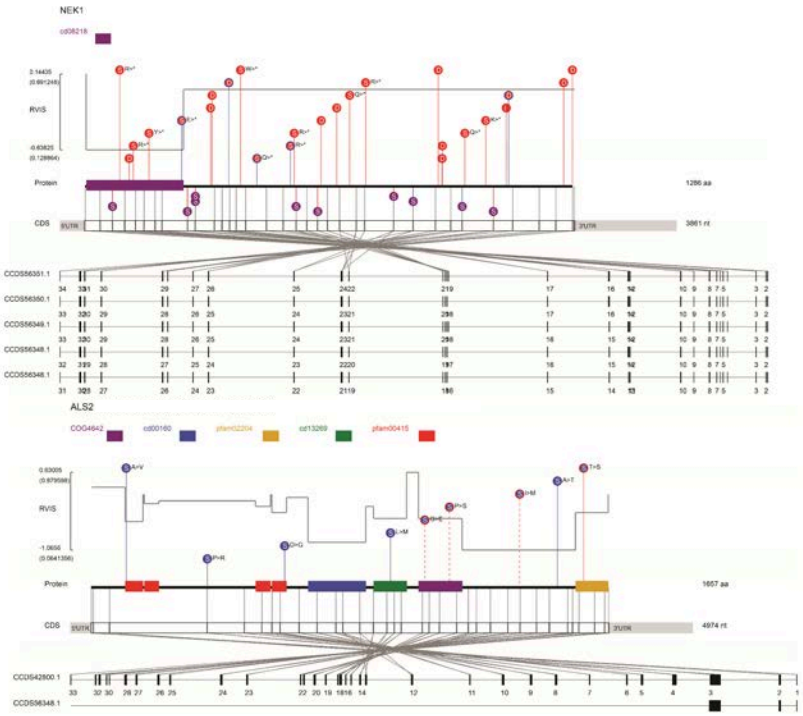

Fig. S3 Variants in *NEK1* and *ALS2*

Dominant LoF variants are shown for *NEK1* (combined dataset), and recessive coding variants are shown for *ALS2* (discovery dataset). LoF variants are filled in red, non-synonymous variants are filled in blue, and splice variants are filled in purple. Case variants are shown with red lines, control variants are shown with blue lines, and variants found in both cases and controls are shown with dashed lines.

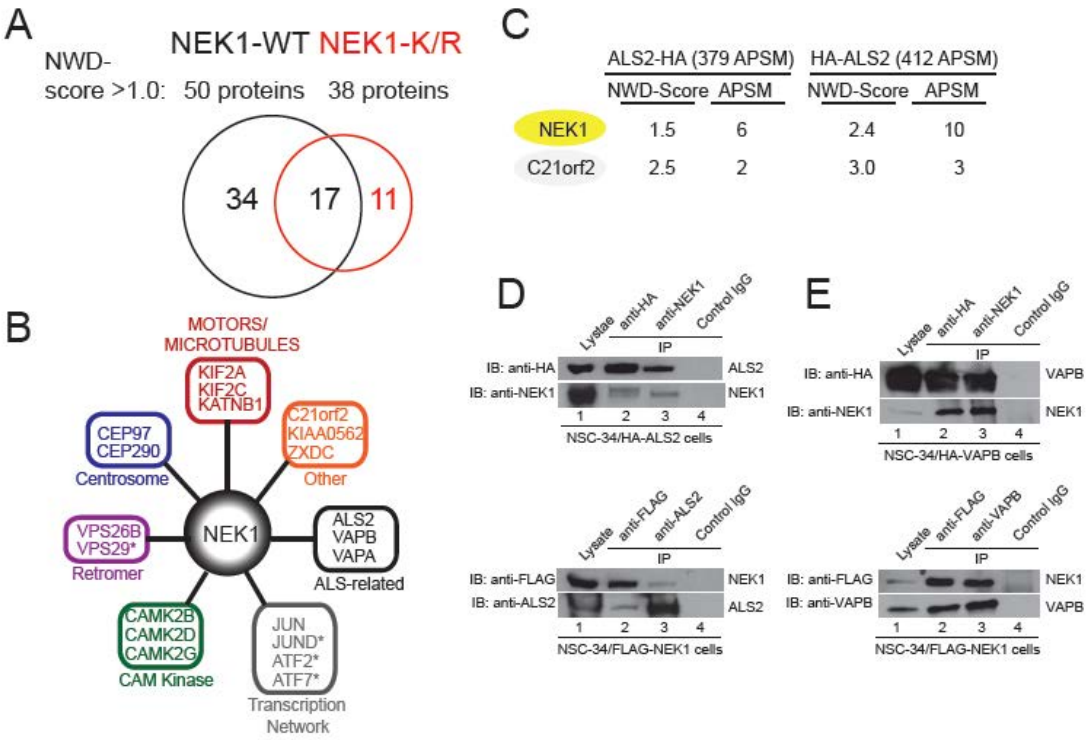

Fig. S4 NEK1 interacts with ALS2 and VAPB

NEK1 interacts with ALS2 and VAPB. (A,B) HEK293T cells stably expressing HA-NEK1 or HA-NEK1^{K33R} (K/R) were subjected to AP-MS analysis using the *CompPASS* platform and proteins with a normalized weighted D (NWD)-score > 1.0 identified. Among the 51 proteins identified with NEK1 and the 38 proteins identified with NEK1^{K33R}, 17 were in common (panel A). The major classes of interacting proteins found with both bait proteins are shown in panel B. Proteins indicated with asterisks were identified but with a sub-threshold NWD-score. (C) HEK293T cells stably expressing either HA-ALS2 or ALS2-HA were subjected to AP-MS and NEK1 as well as the NEK1 associated protein C21orf2 were identified. APSM, average peptide spectral matches. (D,E) NCS-34 neuronal cells expressing either HA-ALS2, HA-VAPB, or FLAG-NEK1 were subjected to immunoprecipitation using either anti-NEK1, anti-VAPB or anti-ALS2, as indicated, to immunoprecipitate the endogenous protein and complexes then immunoblotted with the indicated antibodies. Similarly, anti-FLAG or anti-HA immunoprecipitations were performed to demonstrate reciprocal interactions.

Table S1. Patient Demographics for Discovery Samples

Total ALS Patients Analyzed	2,874
Family History of ALS	6.7% (105/1563)
Male Sex	58.9% (1693/2874)
Bulbar symptom onset	26.8% (661/2467)
If limb onset, proportion upper limb onset	51.2% (909/1774)
Cognitive impairment noted at any time	14.4% (176/1219)
Mean age at symptom onset in years (Stdev, n)	57.1 (13.0,2521)
Range of ages at symptom onset	13-90
Median disease duration in months (IQR, n)	36 (32, 682)

Because data collection varied across centers, the numerators and denominators are shown. Disease duration was only calculated for subjects with complete follow-up and known durations to death or full-time positive pressure ventilation. All patients analyzed were of white ethnicity.

Table S2. Sequencing methods and groups

Kit	37MB	50MB	65MB	Nimblegen	Genome
Analysis group	Cases/Ctrls	Cases/Ctrls	Cases/Ctrls	Cases/Ctrls	Cases/Ctrls
Duke University	0/0	0/0	0/676	1137/2915	36/423
McGill/Stanford University	0/335	248/227	0/165	1/1	2/61

HudsonAlpha	0/0	0/0	0/0	1450/1602	0/0

Numbers are n(%).

Table S3. Number of CCDS bases covered in each analysis

Analysis group	Cases	Ctrls	Cutoff
Duke University	32,233,687 (91%)	32,165,079 (91%)	5%
McGill/Stanford University	28,272,224 (80%)	27,737,938 (78%)	27%
HudsonAlpha	30,969,984 (88%)	31,367,279 (90%)	20%
Replication- Duke University custom capture	111,821	112,423	5%
Replication- University of Massachusetts exomes	102,359	N/A	N/A*

Shown is the average number of bases covered at least 10x. Numbers are n(%). The cutoff refers to the difference allowed between cases and controls in their average exonic coverage; exons with differences above this value were not included in the analysis. *Exomes used in the replication dataset were restricted to the same exons used in the custom capture samples.

Table S4. The 51 genes chosen for targeted follow-up based on initial exome sequencing results.

TET3	*TMEM55B*	*ADCYAP1R1*
ALYREF	*CYGB*	*YBX3*
TAF6	*CYP17A1*	*KCNT1*
NEK1	*CEL*	*CAMK2A*
ATP6V1F	*PCDHGA9*	*LRRC73*
ZNF296	*LGALSL*	*S100A2*
UBE2D2	*PDLIM2*	*HAS2*
DOCK3	*LENG9*	*ZNF837*
TBK1	*C19orf25*	*IL5*
TRAF4	*PAMR1*	*MPL*
OPTN	*SPSB3*	*SLC15A2*
GRID2IP	*DNMT3A*	*PPCS*
C16orf11	*SH3KBP1*	*HIVEP3*
PFKFB1	*SPG11*	*TGM3*
BTBD11	*ZNF432*	*SCEL*
ENAH	*AP1G2*	
TBC1D30	*MADD*	
TNNT3	*GPR162*	

Table S5. NEK1-Interacting proteins

INTERACTOR	NEK1 (APSM)	NEK1K33R (APSM)	NEK1 (NWD)	NEK1K33R (NWD)
ZXDC	10	4	2.75	1.56
VPS29	1	3	0.08	1.35
VPS26B	2	7	1.06	2.19
VAPA	9	4	1.5	1.04
SGPL1	21	4	1.74	0.39
RPS6KA3	4	1	3	0.12
PIPSL	2	0	2.12	0
PDF	1	3	0.12	2.71
NOSIP	2	1	1.06	0.78
NEK1	592	269	17.15	11.95
MYO5A	2	0	1.06	0
MAP4	2	1	1.06	0.78
LRP2	1	2	0.08	1.1
KIFC1	4	1	3	0.12
KIF2C	13	11	3.27	3.06
KIF2A	25	15	2.35	1.72
KIAA0562	15	8	2.77	1.78
KATNB1	2	3	1.06	1.35
KATNA1	4	1	3	0.12
JUN	4	4	1.5	1.56
CREB5	5	0	1.68	0
CEP97	2	1	2.12	1.56
CEP78	5	1	1.12	0.52

CEP290	24	8	3.75	1.78
CAMK2G	4	6	1.5	1.93
CAMK2D	7	10	1.19	1.7
CAMK2B	4	4	1	1.04
C21orf2	10	8	4.74	4.42
ATF7	5	1	3.35	0.12
ATF2	6	1	1.86	0.08
ANKRD27	0	3	0	2.71
ALS2	19	23	2.52	3.01
ALG11	1	0	1.5	0
VPS35	7	13	0.23	0.4
YWHAG	12	13	0.2	0.21
YWHAZ	27	23	0.2	0.19
YWHAQ	22	10	0.15	0.1
YWHAE	104	91	0.48	0.46
YWHAB	28	24	0.46	0.43
YWHAH	10	16	0.36	0.48