NEK1 variants confer susceptibility to amyotrophic lateral sclerosis

Kevin P Kenna[1,40], Perry T C Van Doormaal[2,40], Annelot Dekker[2,40], Nicola Ticozzi[3,4,40], Brendan J Kenna[1], Frank P Diekstra[2], Wouter Van Rheenen[2], Kristel R Van Eijk[2], Ashley R Jones[5], Pamela Keagle[1], Aleksey Shatunov[5], William Sproviero[5], Bradley N Smith[5], Michael A Van Es[2], Simon D Topp[5], Aoife Kenna[1], Jack W Miller[5], Claudia Fallini[1], Cinzia tiloca[3,6], Russell L McLaughlin[7], Caroline Vance[5], Claire Troakes[5], Claudia Colombrita[3,4], Gabriele Mora[8], Andrea Calvo[9], Federico Verde[3,4], Safa Al-Sarraj[5], Andrew King[5], Daniela Calini[3], Jacqueline De Belleroche[10], Frank Baas[11], Anneke J Van Der Kooi[12], Marianne De Visser[12], Anneloor LMA ten Asbroek[11], Peter C Sapp[1], Diane McKenna-Yasek[1], Meraida Polak[13], Seneshaw Asress[13], José Luis Muñoz-Blanco[14], Tim M Strom[15], Thomas Meitinger[16], Karen E Morrison[17], SLAGEN consortium[18], Giuseppe Lauria[19], Kelly L Williams[20], P Nigel Leigh[21], Garth A Nicholson[20,22], Ian P Blair[20], Claire S Leblond[23], Patrick A Dion[23], Guy A Rouleau[23], Hardev Pall[24,25], Pamela J Shaw[26], Martin R Turner[26], Kevin Talbot[26], Franco Taroni[27], Kevin B Boylan[28], Marka Van Blitterswijk[29], Rosa Rademakers[29], Jesús Esteban-Pérez[30,31], Alberto García-Redondo[30,31], Phillip Van Damme[32,33], Wim Robberecht[32,33], Adriano Chio[9], Cinzia Gellera[27], Carsten Drepper[34,35], Michael Sendtner[34], Antonia Ratti[3,4], Jonathan D Glass[13], Jesús S Mora[36], Nazli A Basak[37], Orla Hardiman[7], Albert C Ludolph[38], Peter M Andersen[39], Jochen H Weishaupt[38], Robert H Brown, Jr[1], Ammar Al-Chalabi[5], Vincenzo Silani[3,4,41], Christopher E Shaw[5,41], Leonard H Van Den Berg[2,41], Jan H Veldink[2,41]& John E landers[1,41]

[1]Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [2]Department of Neurology Brain Centre, Brain Centre Rudolf Magnus, University Medical Centre Utrecht, Utrecht, the Netherlands. [3]Department of Neurology, IRCCS Istituto Auxologico Italiano, Milan, Italy. [4]Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center, Università degli Studi di Milano, Milan, Italy. [5]Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, London, UK. [6]Doctoral School in Molecular Medicine, Department of Sciences and Biomedical Technologies, Università degli Studi di Milano, Milan, Italy. [7]Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin, Ireland. [8]Salvatore Maugeri Foundation, IRCSS, Scientific Institute of Milano, Milan, Italy. [9]'Rita Levi Montalcini' Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy. [10]Neurogenetics Group, Division of Brain Sciences, Imperial College London, London, UK. [11]Department of Clinical Genetics, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands. [12]Department of Neurogenetics and Neurology, Academic Medical Centre, University of Amsterdam, Amsterdam, the Netherlands. [13]Department of Neurology, Emory University, Atlanta, Georgia, USA. [14]Unidad de ELA, Instituto de Investigación Hospital Gregorio Marañón de Madrid, Madrid, Spain. [15]Institute of Human Genetics, Helmholtz Zentrum München– German Research Center for Environmental Health, Neuherberg, Germany. [16]Institute of Human Genetics, Technische Universität München, Munich, Germany. [17]Faculty of Medicine, University of Southampton, Southampton, UK. [18]A list of members and affiliations appears at the end of the paper. [19]3rd Neurology Unit, Motor Neuron Diseases Center, Fondazione IRCCS Istituto Neurologica '0sta', Milan, Italy. [20]Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South Wales, Australia. [21]Trafford Centre for Medical Research, Brighton and Sussex Medical School, UK. [22] ANZAC Research Institute, Concord Hospital, University of Sydney, Sydney, New South Wales, Australia. [23]Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Quebec, Canada. [24]Institute of Clinical Studies, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham, UK. [25]Queen Elizabeth Hospital Birmingham, Edgbaston, Birmingham, UK. [26] Nuffield

Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [27]Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologica 'Carlo Besta', Milan, Italy. [28]Department of Neurology, Mayo Clinic Florida, Jacksonville, Florida, USA. [29]Department of Neuroscience, Mayo Clinic, Jacksonville, Florida, USA. [30]Unidad de ELA, Instituto de Investigación Hospital 12 de Octubre de Madrid, Madrid, Spain. [31]Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER) U-723, Madrid, Spain. [32]Laboratory of Neurobiology, Department of Neurosciences, KU Leuven and Vesalius Research Centre, VIB, Leuven, Belgium. [33]Department of Neurology, University Hospitals, Leuven, Belgium. [34]Institute of Clinical Neurobiology, University Hospital Würzburg, Würzburg, Germany. [35]Department of Child and Adolescent Psychiatry, University Hospital of Würzburg, Würzburg, Germany. [36]ALS Unit/Neurology, Hospital San Rafael, Madrid, Spain. [37]NDAL, Department of Molecular Biology and Genetics, Bogazici University, Istanbul, Turkey. [38]Neurology Department, Ulm University, Ulm, Germany. [39]Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå, Sweden. [40]These authors contributed equally to this work. [41]These authors jointly directed this work. Correspondence should be addressed to J.H.V. (j.h.veldink@umcutrecht.nl).

Summary
To identify genetic factors contributing to amyotrophic lateral sclerosis (ALS), we conducted whole exome analyses of 1,022 index familial ALS (FALS) cases and 7,315 controls. In a novel screening strategy, gene burden analyses trained with established ALS genes revealed a significant association between loss of function (LOF) *NEK1* variants and FALS risk. Independently, autozygosity mapping of an isolated community in the Netherlands revealed an *NEK1*:p.R261H variant as a candidate risk factor. Replication analyses of sporadic ALS (SALS) cases and independent control cohorts confirmed significant disease association for both p.R261H (10,589 samples analyzed) and *NEK1* LOF variants (3,362 samples analyzed). In total, *NEK1* risk factors are observed in nearly 3% of ALS cases. *NEK1* has been linked to several cellular functions including cilia formation, DNA damage response, microtubule stability, neuronal morphology and axonal polarity. Our results provide new and important insights into ALS etiopathogenesis and genetic aetiology.

Main Text (1500 words)
In recent years, the combination of exome sequencing, segregation analysis and bioinformatic filtering has proven to be an effective strategy to rapidly identify novel disease genes[1]. Unfortunately, this method can be difficult to apply to disorders such as ALS where late age of onset and low to modest variant penetrance make it difficult to obtain large informative multi-generational pedigrees. Due to high genetic heterogeneity, ALS is also difficult to analyze using filtering methods designed to exploit unrelated patient groups[2]. Recently, we demonstrated the utility of exome-wide rare variant burden analysis (RVB) as an alternative approach, identifying a replicable association between FALS risk and *TUBA4A* in a cohort of 363 cases[3]. In brief, RVB compares the combined frequency of rare variants within each gene in a case-control cohort. Candidate associations are identified by significant differences after multiple test correction. Since this initial study, we have extended our dataset to include complete exome sequencing for 1,376 index FALS cases and 13,883 controls. Of these, 1,022 cases and 7,315 controls met all required data, inter-relatedness and ancestral quality control criteria (Supplementary Fig. 1-2, Online Methods).

The successful detection of disease associations through RVB can depend heavily on the appropriate setting of test parameters. Since genetic loci often contain many alleles of no or low effect, prior filtering of variants based on minor allele frequency (MAF) and pathogenicity

predictors can reveal disease signatures otherwise masked by normal human variability. As appropriate MAF or pathogenicity predictor settings may not be obvious in advance, comprehensive assessment of all pursuable analysis strategies is desirable but can in turn introduce excessive multiple test burden. To overcome these limitations, we performed 308 distinct RVB analyses of 10 well establish ALS genes using 44 functional and 7 MAF filters (Fig. 1a). All tests included correction for gene coverage and ancestral covariates (Online Methods). Within the final cohort, 72 cases and 0 controls harbored known ALS pathogenic mutations within these 10 genes (Online Methods). An additional 26 cases harbored a repeat expansion in the *C9orf72* gene. Tests differed in their capacity to detect individual known ALS genes (Supplementary Table 1), however, highest net sensitivity was achieved when analyses were restricted to variants with MAF<0.001 and functional classifications of either nonsense, splice altering[4] or FATHMM deleterious[5]. Under these settings, 4 genes exhibited disease association at exome-wide (Bonferroni-corrected, $P<2.5\times10^{-6}$) significance (*SOD1, TARDBP, UBQLN2, FUS*), 3 achieved near exome significance (*TUBA4A*, *TBK1, VCP),* and 3 displayed modest to marginal disease association (*PFN1, VAPB, OPTN*) (Fig. 1b). Genes exhibiting the strongest disease associations included those reported as major ALS genes in population based studies while those exhibiting weaker associations are believed to constitute rarer causes of disease.

Extension of the optimal known ALS gene parameters to all protein coding genes revealed one novel gene displaying exome-wide significant disease association (Fig. 1b). The gene, *NEK1* (OR=8.2, $P=1.7\times10^{-6}$), encodes the serine/threonine kinase NIMA (never in mitosis gene-A) related kinase. Retesting of *NEK1* under alternate analysis parameters revealed strong disease associations across most analysis strategies, particularly where loss of function (LOF, nonsense and predicted splice altering) variants were included (Supplementary Table 2). No evidence was observed for systematic genomic inflation (λ=0.95), confounding related to sample ascertainment (Supplementary Fig. 3) or casecontrol biases in *NEK1* gene coverage (Supplementary Fig.4). Removal of samples carrying rare variants of known ALS genes did not influence the association (OR=8.9, $p=7.3\times10^{-7}$).

In an independent line of research, whole genome sequencing was performed for 4 ALS patients from an isolated community in the Netherlands (population<25,000). High inbreeding coefficients were observed for each of the 4 patients confirming their high degree of relatedness and supporting a restricted genetic lineage (Supplementary Fig. 5). Autozygosity mapping, allowing for genetic heterogeneity, identified 4 candidate disease variants occurring within detectable runs of homozygosity (ROH) (Supplementary Fig. 6). These variants included a p.R261H variant of *NEK1*. Two of the 4 SALS cases were homozygous for p.R261H while 2 were heterozygous, raising the possibility that even a single copy of the allele may increase disease risk. Clinical evaluation of the 4 cases did not reveal any overt differences in disease phenotype. . None of the other 3 candidate variants exhibited homozygosity in multiple patients or occurred at all in more than 2 patients. Analysis of the region revealed a shared p.R261H haplotype spanning 3 Mb in all 4 samples (Supplementary Table 3).

To validate the risk effects of p.R261H, we tested for disease association among 6,172 SALS cases and 4,417 matched controls from 8 countries (Supplementary Fig. 7-8, Online Methods). This cohort was either genotyped using the Illumina exome chip or whole genome sequenced, allowing for checking any overlap or detectable relatedness to the FALS casecontrol cohort, which was not present. Meta-analysis of all independent population strata reveal a clear minor allele excess in cases with a combined significance of $p=4.8\times10^{-5}$ and OR=2.4 (Fig. 2). Disease association was also observed within the FALS case-control

data (OR=2.7, p=1.5x10^{-3}) and meta-analysis of FALS, SALS and all controls combined (OR=2.4, *P*=1.2x10^{-7}).

DNA availability facilitated segregation analysis of only one *NEK1* LOF variant, a p.R550X variant which was also detected in the affected mother of the identified proband. To validate the effect of LOF variants observed in FALS and assess any potential contribution to sporadic disease, we analyzed full sequencing data of the *NEK1* coding region for 2,303 SALS and 1,059 controls (Supplementary Fig. 2, Online Methods). RVB confirmed a significant excess of LOF variants in cases (23/2,303 SALS vs 0/1,059 controls, OR=22.2, p=1.5x10^{-4}, Supplementary Table 2). Meta-analysis of discovery and replication LOF analyses yielded a combined significance of *P*=3.4x10^{-8} and OR=8.8.

In total, 120 predicted nonsynonymous *NEK1* variants were detected in FALS, SALS and controls. These were distributed throughout the gene including within the protein kinase domain (PKD) and 6 coiled-coil domains thought to be involved in mediating protein-protein interactions (Supplementary Fig. 9). Following conditioning for LOF and p.R261H, tentative excesses of case variants could be observed in analyses of rarer variant categories but larger sample sizes will be required to confirm the pathogenicity beyond p.R261H and LOF variants (Supplementary Table 4). Analysis of other members of the *NEK* gene family (*NEK2-11*) revealed no associations in the FALS dataset meeting multiple test criteria (Supplementary Table 5).

Although no other gene achieved discovery significance, 10 candidate loci exhibited *P*<1.0x10^{-3} in the FALS discovery analysis (Table 1). These included the gene encoding the SNARE (soluble NSF attachment protein receptor) complex protein synataxin 12 (*STX12*, OR=33.1, *P*=9.7x10^{-5}). Analysis of the SALS replication cohort revealed 5 missense variants in cases vs 0 in controls. However, the cohort was not sufficiently powered to assess events of this frequency and larger sample sizes will be required to establish effects on ALS risk (Supplementary Table 6). Another identified candidate gene was the known hereditary spastic paraplegia gene *KIF5A*[6] (OR=7.1, p=4.8x10^{-4}), however no observed elevations in patient variant frequencies within the SALS replication cohort reached statistical significance (Supplementary Table 7).

NEK1 has been previously described as a candidate gene for ALS[7,8]. Here, our findings reveal that *NEK1* in fact constitutes a major ALS-associated gene with risk variants present in ~3% of European/ European American ALS cases. LOF variants were identified in 1.2% of FALS (OR=8.2) and 1.0% of SALS (OR=22.2) versus 0.17% of controls, while the p.R261H variant was identified in 1.7% of FALS (OR=2.7) and 1.6% of SALS (OR=2.4) versus 0.69% of controls. Other variants of unknown clinical significance (missense, MAF<0.001) were identified in a further 1.8% of FALS and 1.3% of SALS versus 1.2% of controls. In comparison, risk variants in previously established ALS genes occur at approximately the following percentages: *C9orf72*<10%, *SOD1*<2%, *TARDBP*<1%, *FUS*<1% and others <<1% or uncertain[9-12]. However, caution must be taken when comparing the frequency of variants or mutations that differ in penetrance (i.e. highly-penetrant mutations to lower-penetrant risk variants). Furthermore, the assessment of the true odds ratio for variants within a gene may be difficult due to the presence of neutral variants that dilute out the observed effect. The actual odds ratio may in fact be even higher in a specific subset of variants versus controls. The LOF variants within NEK1 display a higher odds ratio relative to p.R261H. The p.R261H variant occurs adjacent to the protein kinase domain and is classified as deleterious by most bioinformatic prediction algorithms (SIFT, PolyPhen, LRT, MutationTaster, Mutation Assessor, PROVEAN, CADD, GERP, SiPhy). One model to account for the difference in

p.R261H and LOF variant toxicity could be a correlation between phenotypic expression and the predicted extent of *NEK1* loss of function. This model would also be consistent with previous findings that homozygosity for *NEK1* LOF variants causes a severe developmental phenotype; short rib polydactyly syndrome type II (SRPS)[13]. In the current study, no individuals carried multiple LOF alleles. However, in SRPS homozygous carriers of *NEK1* LOF variants have been reported to exhibit a 64% reduction of *NEK1* mRNA levels while unaffected heterozygous parents exhibit a 30-40% reduction[13].

NEK1 represents one of 11 members of the highly conserved NIMA-kinase family, which has conserved functions in cell cycle progression and mitosis. In post-mitotic cells, *NEK1* is a primary regulatory of the formation of non-motile primary cilium[14,15]. Disruption in the structure or function of primary cilia have been linked to neurological defects such as brain dysgenesis, hydrocephalus and mental retardation[16,17], and abnormalities in cilia number, structure and microtubule state occur in fibroblasts derived from short rib polydactyly syndrome patients homozygous for *NEK1* truncation variants[13]. *In vitro* disruption of the activity of other neuronally expressed *NEK* family members has similarly been shown to disrupt neuronal morphology, neurite outgrowth, microtubule stability and microtubule dynamics[18,19]. Microtubule organization/integrity and kinesin/dynein intra-flagellar transport are essential to maintain cilia structure and function. This is of particular relevance as disruption of the microtubule cytoskeleton has been associated to the development of ALS[3] and mutations of the dynein subunit dynactin are associated with motor neuron degeneration[20]. Additionally, motor neurons derived from the *SOD1*^{G93A} mouse show a selective loss of cilia both *in vitro* and *in vivo*[21]. Besides its role in ciliogenesis, *NEK1* is also known to regulate mitochondrial membrane permeability[22] and DNA repair[23]. Both these processes have been extensively investigated in relation to ALS, and postulated to explain the toxicity of ALS-associated mutations in *SOD1* and *FUS*[24,25]. Mutations in DNA repair genes cause several early onset neurological phenotypes and multiple lines of evidence suggest defective DNA repair may contribute to both late onset neurodegeneration and brain aging in general[26]. For example, oxidative damage and DNA strand breaks have been observed to be elevated in ALS, Alzheimer's and Parkinson's disease cases[27], and a recent large-scale GWAS implicated DNA repair genes as age of onset modifiers in Huntington's disease[28]. The pathological significance of DNA damage in ALS, and whether modifier effects observed in Huntington's may generalize to repeat expansion disorders like *C9orf72* associated ALS, constitute important questions to be addressed. Finally, through its coiledcoil domain, *NEK1* has been shown to interact with multiple other proteins of potential importance, including the ALS-associated proteins *VAPB* and *ALS2*[7], and the axonal outgrowth regulator *FEZ1*[29].

Data access. Full details of ALS patient variants are publicly available through the ALS Variant Server http://als.umassmed.edu/.

URLs. Exome Variant Server, NHLBI Exome Sequencing Project (ESP), http://evs.gs.washington.edu/EVS/; Exome Aggregation Consortium (ExAC), http://exac.broadinstitute.org; ALS Variant Server http://als.umassmed.edu/.

Accession codes. Genetic Epidemiology of COPD (COPDGene) phs000179; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Cystic Fibrosis) phs000254; 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; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (COPDGene) phs000296; NHLBI Framingham Heart Study Allelic Spectrum Project phs000307; NHLBI GO-ESP: Family Studies (Aortic Disease) phs000347; NHLBI GO-ESP Family Studies: Pulmonary Arterial Hypertension phs000354; NHLBI GO-ESP: Family Studies: (Familial Atrial Fibrillation) phs000362; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (ARIC) phs000398; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (CHS) phs000400; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (FHS) phs000401; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (JHS) phs000402; NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (MESA) phs000403; NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (Asthma) phs000422; Jackson Heart Study Allelic Spectrum Project phs000498; NHLBI GOESP Family Studies: Idiopathic Bronchiectasis phs000518; Alzheimer's Disease Sequencing Project (ADSP) phs000572; NHLBI GO-ESP: Family Studies (Hematological Cancers) phs000632; Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium (CHARGE-S): FHS phs000651; Building on GWAS for NHLBI-Diseases: The U.S. CHARGE Consortium (CHARGE-S): CHS phs000667; Building on GWAS for NHLBI-Diseases : the U.S. CHARGE Consortium (CHARGE-S): ARIC phs000668;. NIH Exome Sequencing of Familial Amyotrophic Lateral Sclerosis Project phs000101.v4.p1.

Acknowledgements
We gratefully acknowledge all of the studies participants and our collaborators for enabling this study by graciously providing samples for this study. Funding was provided by the National Institutes of Health (NIH)/National Institute of Neurological Disorders and Stroke (NINDS) (R01NS073873, JEL), the American ALS Association (NT, VS, CES, JEL, RHB), the MND Association (NT, VS, CES, JEL), the Angel Fund (RHB), Project ALS/P2ALS (RHB), the ALS Therapy Alliance (RHB, JEL). Computer resources for this were provided by the Green High Performance Computing Center at the University of Massachusetts Medical School. PS was supported through the auspices of Dr. H. Robert Horvitz (Massachusetts Institute of Technology), and Investigator of the Howard Hughes Institute. M. van Es received a grant from the Netherlands Organization for Health Research and Development (Veni scheme), travel grants from Baxter, and serves on the biomedical research advisory panel of the UK Motor Neurone Disease Association (MNDA). This is an EU Joint Programme - Neurodegenerative Disease Research (JPND) project.
The project is supported through the following funding organisations under the aegis of JPND - www.jpnd.eu (United Kingdom, Medical Research Council; Netherlands, ZonMW; Italy, Ministero dell'Istruzione, dell'Università e della Ricerca; Belgium, Fonds Wetenschappelijk Onderzoek; Germany, Bundesministerium für Bildung und Forschung). CES and AAC receive salary support from the National Institute for Health Research (NIHR) Dementia Biomedical Research Unit at South London and Maudsley NHS Foundation Trust and King's College London. The work leading up to this publication was funded by the European Community's Health Seventh Framework Programme (FP7/2007–2013; grant agreement number 259867). Samples used in this research were in part obtained from the UK National DNA Bank for MND Research, funded by the MND Association and the Wellcome Trust. We would like to thank people with MND and their families for their participation in this project. I.R.C.C.S. Istituto Auxologico Italiano; AriSLA - Fondazione Italiana di Ricerca per la SLA co-financed with support of "5x1000" - Healthcare Research of the Italian Ministry of Health (grants EXOMEFALS 2009 and NOVALS 2012 [N.T., C.Tiloca, C.G., V.S., J.E.L.]), grant RepeatALS 2013 [S.D., L.C.]), Italian Ministry of Health (grant GR-2011-02347820 - IRisALS [N.T., C.Tiloca, D.C.] This work was supported by a grant from the Flemish agency for Innovation by Science and Technology (IWT, Project MinE), the Interuniversity Attraction Poles (IUAP) program P7/16

of the Belgian Federal Science Policy Office, by the FWO-Vlaanderen under the frame of E-RARE-2, the ERA-Net for Research on Rare Diseases (PYRAMID), by a EU Joint Programme - Neurodegenerative Disease Research (JPND) project (STRENGTH). PVD holds a senior clinical investigatorship of FWO-Vlaanderen and is supported by the Belgian ALS liga. In Australia, this work was supported by a Leadership Grant to IPB from MND Australia and an NHMRC fellowship (1092023) to KLW. G.A.R. is funded by CIHR, Genome-wide exon capture for targeted resequencing in patients with familial amyotrophic lateral sclerosis (#208973) MDA, and Whole exome sequencing in patients with familial ALS (#153959). C.S.L is recipient of Tim E. Noël fellowship from ALS society of Canada. Part of this work was carried out on the Dutch national e-infrastructure with the support of SURF Foundation. The Alzheimer's Disease Sequencing Project (ADSP), phs000572.v7.p4 is comprised of two Alzheimer's Disease (AD) genetics consortia and three National Human Genome Research Institute (NHGRI) funded Large Scale Sequencing Centers (LSSC). The two AD genetics consortia are the Alzheimer's Disease Genetics Consortium (ADGC) funded by NIA (U01 AG032984), and the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) funded by NIA (R01 AG033193), the National Heart, Lung, and Blood Institute (NHLBI), other National Institute of Health (NIH) institutes and other foreign governmental and non-governmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 (to Drs. Schellenberg, Farrer, Pericak-Vance, Mayeux, and Haines; U01AG049505 to Dr. Seshadri; U01AG049506 to Dr. Boerwinkle; U01AG049507 to Dr. Wijsman; and U01AG049508 to Dr. Goate. The ADGC cohorts include: Adult Changes in Thought (ACT), the Alzheimer's Disease Centers (ADC), the Chicago Health and Aging Project (CHAP), the Memory and Aging Project (MAP), Mayo Clinic (MAYO), Mayo Parkinson's Disease controls, University of Miami, the Multi-Institutional Research in Alzheimer's Genetic Epidemiology Study (MIRAGE), the National Cell Repository for Alzheimer's Disease (NCRAD), the National Institute on Aging Late Onset Alzheimer's Disease Family Study (NIA-LOAD), the Religious Orders Study (ROS), the Texas Alzheimer's Research and Care Consortium (TARC), Vanderbilt University/Case Western Reserve University (VAN/CWRU), the Washington Heights-Inwood Columbia Aging Project (WHICAP) and the Washington University Sequencing Project (WUSP), the Columbia University Hispanic- Estudio Familiar de Influencia Genetica de Alzheimer (EFIGA), the University of Toronto (UT), and Genetic Differences (GD). The CHARGE cohorts, with funding provided by 5RC2HL102419 and HL105756, include the following: Atherosclerosis Risk in Communities (ARIC) Study which is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), Austrian Stroke Prevention Study (ASPS), Cardiovascular Health Study (CHS), Erasmus Rucphen Family Study (ERF), Framingham Heart Study (FHS), and Rotterdam Study (RS). The three LSSC are: the Human Genome Sequencing Center at the Baylor College of Medicine (U54 HG003273), the Broad Institute Genome Center (U54HG003067), and the Washington University Genome Institute (U54HG003079). Biological samples and associated phenotypic data used in primary data analyses were stored at Study Investigators institutions, and at the National Cell Repository for Alzheimer's Disease (NCRAD, U24AG021886) at Indiana University funded by NIA. Associated Phenotypic Data used in primary and secondary data analyses were provided by Study Investigators, the NIA funded Alzheimer's Disease Centers (ADCs), and the National Alzheimer's Coordinating Center (NACC, U01AG016976) and the National Institute on Aging Alzheimer's Disease Data Storage Site (NIAGADS, U24AG041689) at the University of Pennsylvania, funded by NIA, and at the Database for Genotypes and Phenotypes

(dbGaP) funded by NIH. Contributors to the Genetic Analysis Data included Study Investigators on projects that were individually funded by NIA, and other NIH institutes, and by private U.S. organizations, or foreign governmental or nongovernmental organizations.

Author Contributions

Sample Collection, Preparation, Clinical Evaluation, Performed Experiments, Data Analysis, Scientific Planning and Direction, Initial Manuscript Preparation Group: KPK, PTCvD, JHV, JEL. Sample Collection, Preparation, Clinical Evaluation, Performed Experiments and Data Analysis Group: AD, NT, BJK, FD, WvR, KRvE, ARJ, PK, BNS, MvE, AK, CF, RLMcL, PCS, DMcK-Y, KLW, CSL, PAD, GAR. Sample Collection, Preparation, Clinical Evaluation, Performed Experiments, Data Analysis, Scientific Planning and Direction Group: IPB, RHB Jr., LHvdB. Sample Collection, Preparation and Clinical Evaluation Group: AS, WS, ST, JMW, CT, CV, C Tr, CC, GM, AC, FV, SA-S, A Ki, DC, JdB, FB, AvdK, MdV, ALMAtA, MP, SA, JLM-B, S D'A, LM, GPC, RDB, MC, SG, GQ, CB, VP, BC, SC, C Cer, LC, GS, GL, GAN, HP, PJS, MRT, KT, FT, KBB, MvB, RR, JE-P, AG-R, PvD, WR, AC, CG, CD, MS, AR, JDG, JSM, ANB, OH, ACL, PMA, JW, AA-C, VS, CES.

Competing financial interests

The authors declare no competing financial interests

1. Gilissen, C., Hoischen, A., Brunner, H.G. & Veltman, J.A. Unlocking Mendelian disease using exome sequencing. *Genome Biol* **12**, 228 (2011).
2. Ng, S.B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-3 (2010).
3. Smith, B.N. *et al.* Exome-wide rare variant analysis identifies TUBA4A mutations associated with familial ALS. *Neuron* **84**, 324-31 (2014).
4. Jian, X., Boerwinkle, E. & Liu, X. In silico prediction of splice-altering single nucleotide variants in the human genome. *Nucleic Acids Res* **42**, 13534-44 (2014).
5. Shihab, H.A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat* **34**, 57-65 (2013).
6. Reid, E. *et al.* A kinesin heavy chain (KIF5A) mutation in hereditary spastic paraplegia (SPG10). *Am J Hum Genet* **71**, 1189-94 (2002).
7. Cirulli, E.T. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347**, 1436-41 (2015).
8. Brenner, D. *et al.* NEK1 mutations in familial amyotrophic lateral sclerosis. *Brain* (2016).
9. Renton, A.E., Chio, A. & Traynor, B.J. State of play in amyotrophic lateral sclerosis genetics. *Nat Neurosci* **17**, 17-23 (2014).
10. Kenna, K.P. *et al.* Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. *J Med Genet* **50**, 776-83 (2013).
11. Lattante, S. *et al.* Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. *Neurology* **79**, 66-72 (2012).
12. Chio, A. *et al.* Extensive genetics of ALS: A population-based study in Italy. *Neurology* **79**, 1983-1989 (2012).

13. Thiel, C. *et al.* NEK1 mutations cause short-rib polydactyly syndrome type majewski. *Am J Hum Genet* **88**, 106-14 (2011).
14. Shalom, O., Shalva, N., Altschuler, Y. & Motro, B. The mammalian Nek1 kinase is involved in primary cilium formation. *FEBS Lett* **582**, 1465-70 (2008).
15. White, M.C. & Quarmby, L.M. The NIMA-family kinase, Nek1 affects the stability of centrosomes and ciliogenesis. *BMC Cell Biol* **9**, 29 (2008).
16. Lee, J.H. & Gleeson, J.G. The role of primary cilia in neuronal function. *Neurobiol Dis* **38**, 167-72 (2010).
17. Lee, L. Riding the wave of ependymal cilia: genetic susceptibility to hydrocephalus in primary ciliary dyskinesia. *J Neurosci Res* **91**, 1117-32 (2013).
18. Cohen, S., Aizer, A., Shav-Tal, Y., Yanai, A. & Motro, B. Nek7 kinase accelerates microtubule dynamic instability. *Biochim Biophys Acta* **1833**, 1104-13 (2013).
19. Chang, J., Baloh, R.H. & Milbrandt, J. The NIMA-family kinase Nek3 regulates microtubule acetylation in neurons. *J Cell Sci* **122**, 2274-82 (2009).
20. Puls, I. *et al.* Mutant dynactin in motor neuron disease. *Nat Genet* **33**, 455-6 (2003).
21. Ma, X., Peterson, R. & Turnbull, J. Adenylyl cyclase type 3, a marker of primary cilia, is reduced in primary cell culture and in lumbar spinal cord in situ in G93A SOD1 mice. *BMC Neurosci* **12**, 71 (2011).
22. Chen, Y., Craigen, W.J. & Riley, D.J. Nek1 regulates cell death and mitochondrial membrane permeability through phosphorylation of VDAC1. *Cell Cycle* **8**, 257-67 (2009).
23. Pelegrini, A.L. *et al.* Nek1 silencing slows down DNA repair and blocks DNA damage-induced cell cycle arrest. *Mutagenesis* **25**, 447-54 (2010).
24. Sama, R.R., Ward, C.L. & Bosco, D.A. Functions of FUS/TLS from DNA repair to stress response: implications for ALS. *ASN Neuro* **6**(2014).
25. Tafuri, F., Ronchi, D., Magri, F., Comi, G.P. & Corti, S. SOD1 misplacing and mitochondrial dysfunction in amyotrophic lateral sclerosis pathogenesis. *Front Cell Neurosci* **9**, 336 (2015).
26. Madabhushi, R., Pan, L. & Tsai, L.H. DNA damage and its links to neurodegeneration. *Neuron* **83**, 266-82 (2014).
27. Coppede, F. & Migliore, L. DNA damage in neurodegenerative diseases. *Mutat Res* **776**, 84-97 (2015).
28. Genetic Modifiers of Huntington's Disease, C. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. *Cell* **162**, 516-26 (2015).
29. Surpili, M.J., Delben, T.M. & Kobarg, J. Identification of proteins that interact with the central coiled-coil region of the human protein kinase NEK1. *Biochemistry* **42**, 15369-76 (2003).

SLAGEN Consortium

Sandra d'Alfonso[42], Letizia Mazzini[43], Giacomo P Comi[4,44], Roberto Del Bo[4,44], Mauro Ceroni[45,46], Stella Gagliardi[45], Giorgia Querin[47], Cinzia Bertolin[47], Viviana Pensato[27], Barbara Castellotti[27], Stefania Corti[4,44], Cristina Cereda[45], Lucia Corrado[42] & Gianni Sorarù[4742]
Department of Health Sciences, University of Eastern Piedmont, Novara, Italy. [43]ALS Center, Department of Neurology, 'A. Avogadro' University of Eastern Piedmont, Novara, Italy.

[44]Neurology Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [45]Experimental Neurobiology Laboratory, 'C. Mondino' National Institute of Neurology Foundation, IRCCS, Pavia, Italy. [46]Department of Neurological Sciences, University of Pavia, Pavia, Italy. [47]Department of Neurosciences, University of Padova, Padova, Italy.

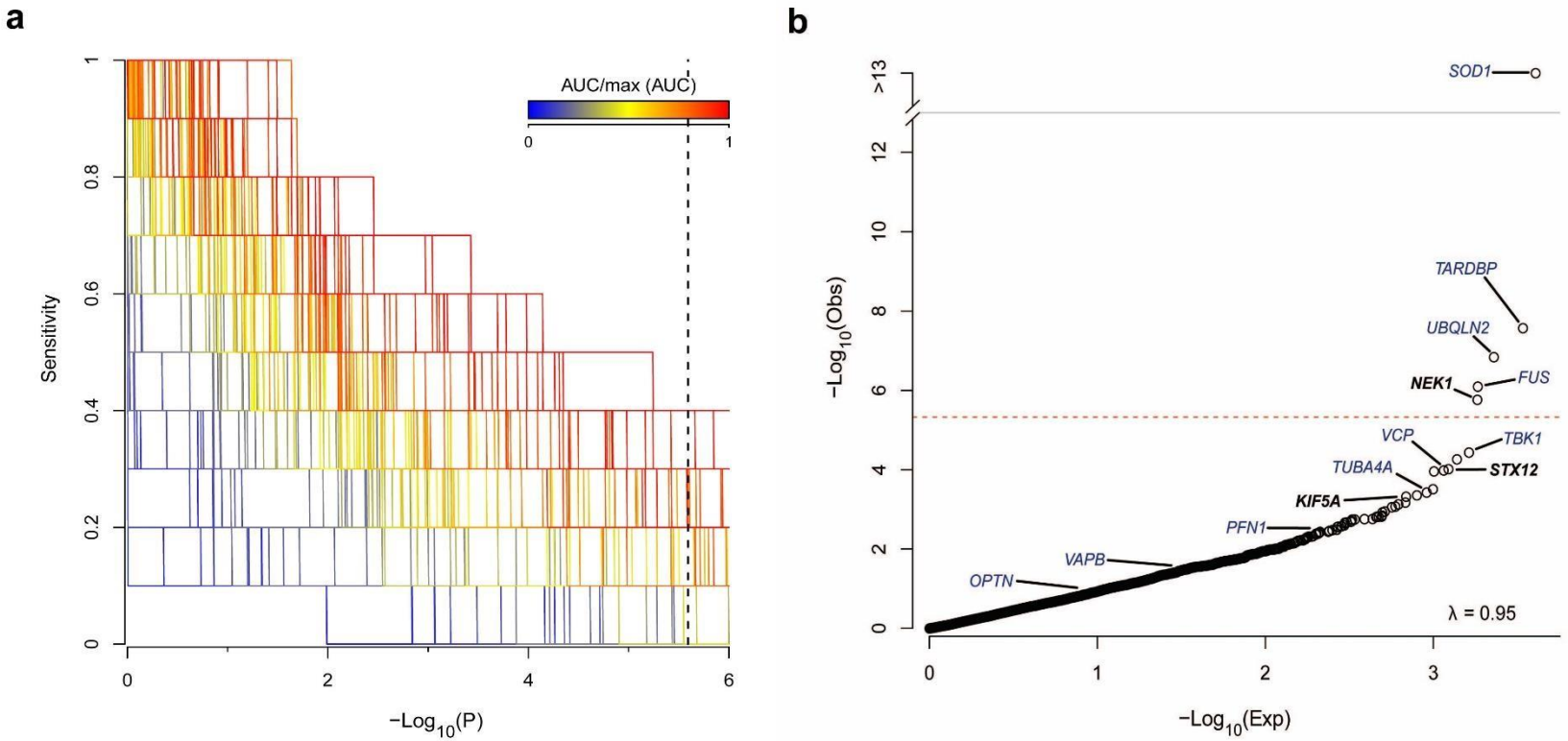

Figure 1: Rare Variant Burden Analysis of FALS Exomes. (a) Rare variant burden analyses of 1,022 index FALS cases and 7,315 controls were performed for 10 known ALS genes. Analyses assessed 308 different combinations of minor allele frequency and functional prediction filters (Supplementary Table 1). The set of analysis parameters achieving the highest sensitivity for known ALS genes was identified as that achieving the highest area under the curve (AUC) in a plot of sensitivity (proportion of training genes achieving significance) across an increasing minimum p-value threshold. Dotted vertical line denotes Bonferroni corrected p-value for exome-wide significance. (b) Extension of the highest performing known gene trained analysis to the entire exome. Threshold for exome-wide significance is denoted by the dotted red line.

Cohort	Cases	Controls	MAF (Cases)	MAF (Controls)	OR	95% CI	*P*
FALS	**1,022**	**7,315**	**0.0086**	**0.0036**	**2.66**	**1.48−4.57**	**1.5×10^{-3}**
SALS							
Belgium	466	476	0.0097	0.0053	1.81	0.60−5.51	
Spain / Italy	472	183	0.0074	0.0055	1.36	0.28−6.58	
Germany	1,229	288	0.0090	0.0017	5.27	0.71−38.81	
Ireland	565	526	0.0044	0.0019	2.32	0.45−12.05	
Netherlands	1,839	1,982	0.0109	0.0035	2.99	1.63-5.47	
United Kingdom	1,335	893	0.0049	0.0022	2.59	0.71-6.69	
United States	266	69	0.0056	0.0072	0.84	0.08-8.29	
Total	**6,172**	**4,417**	**0.0080**	**0.0033**	**2.41**	**1.58−3.71**	**4.8×10^{-5}**
FALS + SALS	**7,194**	**11,732**	**0.0081**	**0.0035**	**2.41**	**1.57−3.71**	**1.2×10^{-7}**

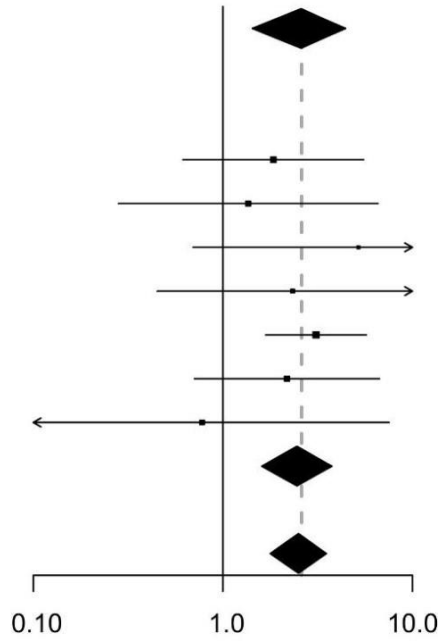

Figure 2: Replication analysis of *NEK1*:p.R261H. *NEK1*:p.R261H genotypes were ascertained for 1,022 FALS, 6,172 SALS and 11,732 controls. The SALS cohort was divided into 7 geographically based case-control strata. Allelic tests of association were performed for all subcohorts and followed by meta-analysis.

Gene	ALS	ALS Freq	Control	Control Freq	OR	OR 95% CI	P
NEK1	12	0.0117	14	0.0019	8.2	3.7-18.0	$1.7x10^{-6}$
ATRN	8	0.0078	7	0.0010	10.3	3.6-29.6	$3.7x10^{-5}$
STX12	4	0.0039	1	0.0001	33.1	5.8-339.0	$9.7x10^{-5}$
CREB3L2	4	0.0039	0	0.0000	64.9	6.6-8695.3	$1.1x10^{-4}$
DCC	4	0.0039	2	0.0003	18.6	4.1-108.1	$3.1x10^{-4}$
WDR49	5	0.0049	2	0.0003	15.8	3.5-92.1	$4.4x10^{-4}$
KIF5A	7	0.0068	8	0.0011	7.1	2.5-19.7	$4.8x10^{-4}$
C1QTNF7	12	0.0117	26	0.0036	3.6	1.8-7.1	$6.7x10^{-4}$
PEAK1	5	0.0049	3	0.0004	11.6	2.9-51.5	$7.5x10^{-4}$
BIRC6	10	0.0098	18	0.0025	4.3	1.9-9.3	$8.4x10^{-4}$
ZSCAN5B	4	0.0039	2	0.0003	16.3	3.3-98.0	$8.8x10^{-4}$

Table 1: FALS discovery analysis candidate genes. RVB results for all genes exhibiting case association at $P<1x10^{-3}$ in FALS discovery cohort.

Online Methods

FALS discovery cohort. The FALS discovery cohort included 1,376 FALS patients and 13,883 non-ALS controls analyzed by exome sequencing. Patients were recruited at specialist clinics in Ireland (n=18), Italy (n=143), Spain (n=49), the UK (n=219), the USA (n=511), Netherlands (n=50), Canada (n=34), Belgium (n=12), Germany (n=202), Turkey (n=47) and Australia (n=91). Variants occurring at very low frequency in the general population (ExAC MAF<0.0001) which have been both previously reported as ALS associated and annotated as either 'pathogenic' or 'likely pathogenic' by Clinvar within the 10 genes were considered to be pathogenic mutations. The breakdown of the 72 mutations observed in the final cohort included the following: *SOD1* (28), *TARDBP* (12), *FUS* (9), *PFN1* (6), *TBK1* (1), *TUBA4A* (4), *UBQLN2* (4), *VAPB* (2), *VCP* (6). An additional 26 cases harbored a repeat expansion in the *C9ORF72* gene. Controls included 29 internal samples and individuals participating in the dbGAP[30] projects listed under accession codes.. Familial history was considered positive for ALS if the proband had at least one affected relative within three generations. 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).

SALS replication cohort: The SALS replication cohort included 2,387 SALS and 1,093 controls analyzed by whole genome sequencing and 5,834 SALS and 4,117 controls analyzed by exome chip. All individuals were recruited at specialist clinics in Ireland, Italy, Spain, the UK, the USA, the Netherlands and Belgium. Details of sample contributions per country are shown in Fig. 2. Evaluation of *C9orf72* status was performed in 2,387 SALS cases and 166 (7%) displayed a repeat expansion. 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).

Exome sequencing. Exome sequencing of patients was performed as previously described[3]. Raw sequence data for controls was obtained from dbGAP. Sequence reads were aligned to human reference GRCh37 using BWA (Burrows-Wheeler Aligner) and processed according to recommended best practices[31]. Variant detection and genotyping were performed using the GATK HaplotypeCaller. Variant quality control was performed using the GATK variant quality score recalibration method, with a VQSLOD cut-off of 2.27 (truth set sensitivity of 99%). 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). Exome sequencing data was not used to infer the presence or absence of indels due to the limited sensitivity and comparatively high false positive rates associated with available calling algorithms[32].

Genome Sequencing: Whole genome sequencing of 2387 SALS and 1093 controls was performed with Illumina's FastTrack services (San Diego, USA) using PCR free library preparation and paired-end (100bp or 150bp) sequencing on the HiSeq 2500 or Hiseq X platform (Illumina®, San Diego, Illumina) to yield 35X coverage at minimum. BWA was used to align sequencing reads to genome build hg19, and the Isaac variant caller was used to call single nucleotide variants (SNVs), insertions and deletions (indels)[33]. Both the aligned and unaligned reads were delivered in binary sequence alignment/map format (BAM) together with variant call format (VCF) files containing the SNVs, and indels. gVCF files were generated per individual and variants that failed the Isaac-based quality filter were excluded.

Exome chip: A total of 5,815 ALS patients and 4,614 healthy controls from the Netherlands, Belgium, Germany, Ireland, Italy, Spain and the UK were included. Genotyping was conducted using Illumina HumanExome-12v1 BeadChips in accordance with the manufacturer's recommendations. The GenTrain 2.0 clustering algorithm was used for genotype calling, as implemented in the Illumina GenomeStudio software package. Initial genotype calls were made based on the HumanExome clusterfile provided by Illumina. More accurate cluster boundaries were determined based on the actual study data, after the exclusion of samples with a GenCall quality score in the lower 10th percentile of the distribution across all variants genotyped

(p10GC) < 0.38 or call rate <0.99. Subsequently, the excluded samples were added back into the data set, and new genotypes calls were made using the previously obtained cluster boundaries.

Sample filtering: Samples from the FALS discovery and SALS replication cohorts were excluded from analysis in the event of failing to meet genotype call rate, heterozygosity, gender concordance, duplication, relatedness or population stratification filters as summarized in Supplemental Fig. 1 and Supplemental Fig 7. All samples from the FALS cohort were required to exhibit filtered exome-wide call rates >70%. For both the FALS and SALS cohorts, PLINK v1.07[34] was used to define an LD pruned (R^2<0.5, window size=50, step=5) set of autosomal markers with MAF>0.01 and p>0.001 for deviation from Hardy Weinberg equilibrium. These marker sets were then used to calculate inbreeding coefficients for use in heterozygosity filtering, identify study duplicates, conduct relatedness filtering, perform tests of pairwise population concordance for stratification filtering, conduct principal components (PC) analysis for a second round of stratification filtering and conduct PC analysis to generate covariates for stratification correction in RVB and single variant analysis of filtered cohorts. Samples from the SALS replication cohort were required to exhibit no relatedness/ duplication with samples from the FALS discovery cohort. PLINK was used to calculate inbreeding coefficients, test for discordance in reported and SNV predicted gender and conduct tests of pairwise population concordance. Identification of sample duplicates and sample relatedness was performed using KING[35]. PC analyses were conducted using GCTA[36] (Genome-wide Complex Trait Analysis). Details of results from population stratification analysis provided in Supplementary Fig. 2 and Supplementary Fig. 8.

Statistical analyses: RVB analyses were performed by logistic regression of case-control status to number of minor alleles observed per sample per gene[3,37]. Results from underpowered tests ($\leq$3 observations in combined case control cohort) were excluded and did not contribute to assessments of genomic inflation. Variants were included for RVB analyses on the basis of MAF within the combined case-control cohort, MAF within the 1000 genomes project[38], and pathogenicity predictions generated using snpEFF[39] (Single Nucleotide Polymorphism Effect), PolyPhen2[40] (Polymorphism Phenotyping version2), SIFT[41] (Sorting Intolerant From Tolerant), LRT[42] (Likelihood Ratio Test), MutationTaster[43], MutationAssessor[44], FATHMM[5] (Functional Analysis through Hidden Markov Models), CADD[45] (Combined Annotation Dependent Depletion), PROVEAN[46] (Protein Variation Effect Analyzer), GERP[47] (Genomic Evolutionary Rate Profiling), phyloP[48] (Phylogenetic P-value), SiPhy[49] (SiPhylogenic), dbNSFP[50] (database Nonsynonymous SNP Functional Prediction) and dbscSNV[4] (database of splice site consequences of Single Nucleotide Variants) as described Supplemental table 1. All RVB analyses were conditioned for a missing variant MAF weighted measure of sample gene call rate and the first 4 PC derived from common variant profiles. Homozygosity mapping was performed using HomozygosityMapper[51] allowing for genetic heterogeneity. ROH were selected as all loci achieving a homozygosity score>=8483 (0.6 x max). Single variant analyses were allele count based, conducted using PLINK and also included correction for the first 4 PC derived from common variant profiles. Meta-analyses were conducted using METAL[52] under a fixed effect model with weighting by inverted effect size standard error. All statistical tests were two-sided.

30. Tryka, K.A. *et al.* NCBI's Database of Genotypes and Phenotypes: dbGaP. *Nucleic Acids Res* **42**, D975-9 (2014).
31. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
32. Fang, H. *et al.* Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med* **6**, 89 (2014).
33. Raczy, C. *et al.* Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* **29**, 2041-3 (2013).
34. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).

35. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
36. Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011).
37. Heinze, G. & Schemper, M. A solution to the problem of separation in logistic regression. *Stat Med* **21**, 2409-19 (2002).
38. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
39. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80-92 (2012).
40. Adzhubei, I., Jordan, D.M. & Sunyaev, S.R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet* **Chapter 7**, Unit7 20 (2013).
41. Kumar, P., Henikoff, S. & Ng, P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**, 1073-81 (2009).
42. Chun, S. & Fay, J.C. Identification of deleterious mutations within three human genomes. *Genome Res* **19**, 1553-61 (2009).
43. Schwarz, J.M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods* **7**, 575-6 (2010).
44. Reva, B., Antipin, Y. & Sander, C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic Acids Res* **39**, e118 (2011).
45. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**, 310-5 (2014).
46. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R. & Chan, A.P. Predicting the functional effect of amino acid substitutions and indels. *PLoS One* **7**, e46688 (2012).
47. Davydov, E.V. *et al.* Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol* **6**, e1001025 (2010).
48. Cooper, G.M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**, 901-13 (2005).
49. Garber, M. *et al.* Identifying novel constrained elements by exploiting biased substitution patterns. *Bioinformatics* **25**, i54-62 (2009).
50. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Non-synonymous and Splice Site SNVs. *Hum Mutat* (2015).
51. Seelow, D., Schuelke, M., Hildebrandt, F. & Nurnberg, P. HomozygosityMapper--an interactive approach to homozygosity mapping. *Nucleic Acids Res* **37**, W593-9 (2009).
52. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).