Published in final edited form as:

Nat Genet. 2016 September ; 48(9): 1037–1042. doi:10.1038/ng.3626.

NEK1 variants confer susceptibility to amyotrophic lateral sclerosis

A full list of authors and affiliations appears at the end of the article.

Abstract

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 new screening strategy, we performed gene-burden analyses trained with established ALS genes and identified a significant association between loss-of-function (LOF) *NEK1* variants and FALS risk. Independently, autozygosity mapping for an isolated community in the Netherlands identified a *NEK1* p.Arg261His variant as a candidate risk factor. Replication analyses of sporadic ALS (SALS) cases and independent control cohorts confirmed significant disease association for both p.Arg261His (10,589 samples analyzed) and *NEK1* LOF variants (3,362 samples analyzed). In

Correspondence should be addressed to J.H.V. (j.h.veldink@umcutrecht.nl).

[18]A list of members and affiliations appears at the end of the paper

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[47]

[42]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.

[40]These authors contributed equally to this work

[41]These authors jointly directed this work

Data access. Full details of variants identified in ALS patients are publicly available through the ALS Variant Server at 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/.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS

Sample collection, preparation and clinical evaluation: P.T.C.v.D., A.M.D., N.T., F.P.D., W.v.R., K.R.v.E., A.R.J., P.K., A.S., W.S., B.N.S., M.A.v.E., S.D.T., A. Kenna, J.W.M., C. Tiloca, R.L.M., C.V., C. Troakes, C. Colombrita, G.M., A. Calvo, F.V., S.A.-S., A. King, D.C., J.d.B., F.B., A.J.v.d.K., M.d.V., A.L.M.A.t.A., P.C.S., D.M.-Y., M.P., S.A., J.L.M.-B., T.M.S., T.M., K.E.M., S.D'A., L.M., G.P.C., R.D.B., M.C., S.G., G.Q., C.B., V.P., B.C., S.C., C. Cereda, L.C., G.S., G.L., K.L.W., P.N.L., G.A.N., I.P.B., C.S.L., P.A.D., G.A.R., H.P., P.J.S., M.R.T., K.T., F.T., K.B.B., M.V.B., R.R., J.E.-P., A.G.-R., P.V.D., W.R., A. Chio, C.G., C.D., M.S., A.R., J.D.G., J.S.M., N.A.B., O.H., A.C.L., P.M.A., J.H.W., R.H.B., A.A.-C., V.S., C.E.S., L.H.v.d.B., J.H.V. and J.E.L. Experiments and data analysis: K.P.K., P.T.C.v.D., A.M.D., N.T., B.J.K., F.P.D., W.v.R., K.R.v.E., A.R.J., P.K., A.S., W.S., B.N.S., M.A.v.E., S.D.T., A. Kenna, J.W.M., C.F., C.T., R.L.M., C.V., C. Troakes, C. Colombrita, G.M., A. Calvo, F.V., S.A.-S., A. King, D.C., P.C.S., D.M.-Y., K.L.W., C.S.L., P.A.D., M.v.B., R.R., J.E.-P., A.G.-R., P.v.D., W.R., A. Chio, C.G., C.D., M.S., A.R., J.D.G., J.S.M., N.A.B., O.H., A.C.L., P.M.A., J.H.W., R.H.B.Jr, A.A.-C., V.S., C.E.S., L.H.v.d.B., J.H.V. and J.E.L. Scientific planning and direction: K.P.K., P.T.C.v.D., A.M.D., N.T., B.J.K., C.F., I.P.B., C.S.L., P.A.D., G.A.R., H.P., P.J.S., M.R.T., K.T., F.T., K.B.B., M.v.B., R.R., J.E.-P., A.G.-R., P.v.D., W.R., A. Chio, C.G., C.D., M.S., A.R., J.D.G., J.S.M., N.A.B., O.H., A.C.L., P.M.A., J.H.W., R.H.B.Jr, A.A.-C., V.S., C.E.S., L.H.v.d.B., J.H.V. and J.E.L. Initial manuscript preparation: K.P.K., P.T.C.v.D., A.M.D., N.T., A.A.-C., V.S., C.E.S., L.H.v.d.B., J.H.V. and J.E.L.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

total, we observed *NEK1* risk variants 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 etiology.

In recent years, the combination of exome sequencing, segregation analysis and bioinformatic filtering has proven to be an effective strategy to rapidly identify new disease genes[1]. Unfortunately, this method can be difficult to apply to disorders such as ALS, for which late age of onset and low-to-modest variant penetrance make it difficult to obtain large informative multigenerational pedigrees. Owing to high genetic heterogeneity, ALS is also difficult to analyze using filtering methods designed to exploit unrelated patient groups[2]. Recently, we had demonstrated the utility of exome-wide rare variant burden (RVB) analysis as an alternate approach, identifying a replicable association between FALS risk and *TUBA4A* in a cohort of 363 cases[3]. In brief, RVB analysis is used to compare the combined frequency of rare variants in each gene in a case–control cohort. Candidate associations are identified by significant differences after multiple-test correction. Since this initial study, we extended our data set 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 Figs. 1 and 2, and Online Methods).

Successful detection of disease associations through RVB analysis can depend heavily on the appropriate setting of test parameters. As 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 identify 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 ten well-established ALS genes using 44 functional and 7 MAF filters (Fig. 1a). All tests included correction for gene coverage and ancestral covariates (Online Methods). In the final cohort, 72 cases and 0 controls harbored known ALS pathogenic mutations in these ten 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), but we achieved the highest net sensitivity when we restricted analyses to variants with MAF < 0.001 and functional classifications of either nonsense, splice-altering[4] or deemed deleterious by functional analysis through hidden Markov models (FATHMM)[5]. Under these settings, four genes exhibited disease association at exome-wide (Bonferroni-corrected $P < 2.5 \times 10^{-6}$) significance (*SOD1*, *TARDBP*, *UBQLN2* and *FUS*), three achieved near exome-wide significance (*TUBA4A*, *TBK1* and *VCP*), and three displayed modest to marginal disease association (*PFN1*, *VAPB* and *OPTN*) (Fig. 1b). Genes exhibiting the strongest disease associations included those reported as major ALS genes in population-based studies, whereas 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 identified one new gene displaying exome-wide significant disease association (Fig. 1b). The gene, *NEK1* (odds ratio (OR) = 8.2, $P = 1.7 \times 10^{-6}$), encodes the serine/threonine kinase NIMA (never in mitosis gene-A)-related kinase. Retesting *NEK1* under alternate analysis parameters identified strong disease associations across most analysis strategies, particularly where we included LOF (nonsense and predicted splice-altering) variants (Supplementary Table 2 and Supplementary Fig. 3). We observed no evidence for systematic genomic inflation ($\lambda = 0.95$), confounding related to sample ascertainment (Supplementary Fig. 4) or case–control biases in *NEK1* gene coverage (Supplementary Fig. 5). Removal of samples carrying rare variants of known ALS genes did not influence the association (OR = 8.9, $P = 7.3 \times 10^{-7}$).

In an independent line of research, we performed whole-genome sequencing for four ALS patients from an isolated community in the Netherlands (population < 25,000). We observed high inbreeding coefficients for each of the four patients, confirming their high degree of relatedness and supporting a restricted genetic lineage (Supplementary Fig. 6). Autozygosity mapping, allowing for genetic heterogeneity, identified four candidate disease variants occurring in detectable runs of homozygosity (ROH) (Supplementary Fig. 7). These variants included a p.Arg261His variant of *NEK1*. Two of the four SALS cases were homozygous for p.Arg261His and two were heterozygous, raising the possibility that even a single copy of the allele may increase disease risk. Clinical evaluation of the four cases did not find any overt differences in disease phenotype. None of the other three candidate variants exhibited homozygosity in multiple patients or occurred at all in more than two patients. Analysis of the region identified a shared p.Arg261His haplotype spanning 3 Mb in all four samples (Supplementary Table 3).

To validate the risk effects of p.Arg261His, we tested for disease association among 6,172 SALS cases and 4,417 matched controls from eight countries (Supplementary Figs. 8 and 9, and Online Methods). We genotyped this cohort using the Illumina exome chip or by whole-genome sequencing, allowing for checking of any overlap or detectable relatedness to the FALS case–control cohort, which was not present. Meta-analysis of all independent population strata identified a clear minor allele excess in cases with a combined significance of $P = 4.8 \times 10^{-5}$ and OR = 2.4 (Fig. 2). We also observed disease association in the FALS case–control data (OR = 2.7, $P = 1.5 \times 10^{-3}$) and a meta-analysis of FALS, SALS and all controls combined (OR = 2.4, $P = 1.2 \times 10^{-7}$).

DNA availability facilitated segregation analysis of only one *NEK1* LOF variant, a p.Arg550* variant, which we 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 cases and 1,059 controls (Supplementary Fig. 3 and Online Methods). RVB analysis confirmed a significant excess of LOF variants in cases (23/2,303 SALS samples versus 0/1,059 controls, OR = 22.2, $P = 1.5 \times 10^{-4}$; Supplementary Table 2). Meta-analysis of discovery and replication LOF analyses yielded a combined significance of $P = 3.4 \times 10^{-8}$ and OR = 8.8.

In total, we detected 120 predicted nonsynonymous *NEK1* variants in FALS samples, SALS samples and controls. These were distributed throughout the gene including in the sequence encoding protein kinase domain (PKD) and six coiled-coil domains thought to be involved in mediating protein–protein interactions (Supplementary Fig. 3). After conditioning for LOF variants and p.Arg261His, we observed tentative excesses of case variants in analyses of rarer variant categories, but larger sample sizes will be required to confirm the pathogenicity beyond p.Arg261His and LOF variants (Supplementary Table 4). Analysis of other members of the *NEK* gene family (*NEK2*–*NEK11*) identified no associations in the FALS data set meeting multiple-test criteria (Supplementary Table 5).

Although no other gene achieved discovery significance, ten candidate loci exhibited $P < 1.0 \times 10^{-3}$ 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, $P = 9.7 \times 10^{-5}$). Analysis of the SALS replication cohort identified missense variants in 5/2,303 cases versus 0/1,059 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, $P = 4.8 \times 10^{-4}$); 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 show that *NEK1* in fact constitutes a major ALS-associated gene with risk variants present in ~3% of European and European-American ALS cases. We identified LOF variants in 1.2% of FALS samples (OR = 8.2) and 1.0% of SALS samples (OR = 22.2) versus 0.17% of controls, whereas we identified the p.Arg261His variant in 1.7% of FALS samples (OR = 2.7) and 1.6% of SALS samples (OR = 2.4) versus 0.69% of controls. We identified variants of unknown clinical importance (missense, MAF < 0.001) in a further 1.8% of FALS samples and 1.3% of SALS samples 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-penetrance risk variants). Furthermore, assessment of the true odds ratio for variants in a gene may be difficult because of the presence of neutral variants that dilute out the observed effect. The actual odds ratio may therefore be even higher for specific subsets of patient variants. The LOF variants in *NEK1* displayed a higher odds ratio relative to p.Arg261His. The p.Arg261His 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 and SiPhy). One model to account for the difference in p.Arg261His and LOF variant toxicity could be a correlation between phenotypic expression and the predicted extent of *NEK1* LOF. 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 whereas 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 postmitotic cells, *NEK1* is a primary regulator of the formation of nonmotile primary cilium[14,15]. Disruption in the structure or function of primary cilia has been linked to neurological defects such as brain dysgenesis, hydrocephalus and intellectual disability[16,17], and abnormalities in cilia number, structure and microtubule state occur in fibroblasts derived from SRPS 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 integrity and kinesin and dynein intraflagellar 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 mice expressing human SOD1 G93A 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 of these processes have been extensively investigated in relation to ALS, and have been 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 disease and Parkinson's disease cases[27], and a recent large-scale genome-wide association study (GWAS) implicated DNA-repair genes as age-of-onset modifiers in Huntington's disease[28]. The pathological importance of DNA damage in ALS, and whether modifier effects observed in Huntington's disease may generalize to repeat-expansion disorders such as *C9orf72*-associated ALS, constitute important questions to be addressed. Finally, through its coiled-coil domain, NEK1 has been shown to interact with multiple other proteins of potential importance, including the ALS-associated proteins VAPB and ALS2 (ref. [7]) and the axonal outgrowth regulator FEZ1 (ref. [29]).

METHODS

Methods and any associated references are available in the online version of the paper.

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 United States (n = 511), the 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 ten 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) and *VCP* (6). An additional 26 cases harbored a repeat expansion in the *C9orf72* gene. Controls included 29 internal samples and samples obtained from dbGAP[30]. Sequencing obtained from dbGAP was generated under the following projects: Genetic Epidemiology of chronic obstructive pulmonary disease (COPD) (COPDGene) phs000179; NHLBI Grand Opportunity Exome Sequencing Project (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 (Thoracic aortic aneurysms leading to acute aortic dissections) 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 GO-ESP 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 US CHARGE consortium (CHARGE-S): FHS phs000651; Building on GWAS for NHLBI diseases: The US CHARGE Consortium (CHARGE-S): CHS phs000667; Building on GWAS for NHLBI Diseases: the US CHARGE Consortium (CHARGE-S): ARIC phs000668; NIH Exome Sequencing of FALS Project phs000101.v4.p1. 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 cases and 1,093 controls analyzed by whole-genome sequencing, and 5,834 SALS cases and 4,117 controls analyzed by exome chip. All individuals were recruited at specialist clinics in Ireland, Italy, Spain, the UK, the United States, the Netherlands and Belgium. Details of sample contributions per country are shown in Figure 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 Burrows–Wheeler aligner (BWA) 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 cutoff 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 2,387 SALS samples and 1,093 controls was performed with Illumina's FastTrack services using PCR free library preparation and paired-end (100 bp or 150 bp) sequencing on the HiSeq 2500 or Hiseq X platform (Illumina) to yield 35× 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 10^{th} 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 Supplementary Figures 1 and 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 PCA for a second round of stratification filtering and conduct PCA to generate covariates for stratification correction in RVB analysis 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]. PCA was conducted using genome-wide complex trait analysis (GCTA)[36]. Details of results from population stratification analysis are provided in Supplementary Figures 2 and 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 (≤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 (single nucleotide polymorphism effect)[39], PolyPhen2 (polymorphism phenotyping version 2)[40], SIFT (sorting intolerant from tolerant)[41], LRT (likelihood ratio test)[42], MutationTaster[43], MutationAssessor[44], FATHMM (functional analysis through hidden Markov models)[5], CADD (combined annotation dependent depletion)[45], PROVEAN (protein variation effect analyzer)[46], GERP (genomic evolutionary rate profiling)[47], phyloP (phylogenetic P value)[48], SiPhy (SiPhylogenic)[49], dbNSFP (database nonsynonymous SNP functional prediction)[50] and dbscSNV (database of splice site consequences of single nucleotide variants)[4] as described in Supplementary Table 1. All RVB analyses were conditioned for a missing variant MAF-weighted measure of sample gene call rate and the first four PCs 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 ≥ 8,483 (0.6 × max). Single variant analyses were allele-count-based, conducted using PLINK, and also included correction for the first four PCs 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Kevin P Kenna[1,40], Perry T C van Doormaal[2,40], Annelot M 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 L M A 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], and John E Landers[1,41]

Affiliations

[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 [19]3rd Neurology Unit, Motor Neuron Diseases Center, Fondazione IRCCS Istituto Neurologico 'Carlo Besta',

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, Falmer, 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]Department of Neurology, 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 Neurologico '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

Acknowledgments

We acknowledge all of the study participants and our collaborators for enabling this study by graciously providing samples for this study. Funding was provided by US National Institutes of Health (NIH)/National Institute of Neurological Disorders and Stroke (NINDS) (R01NS073873, J.E.L.), the American ALS Association (N.T., V.S., C.E.S., J.E.L. and R.H.B.Jr.), the Motor Neuron Disease (MND) Association (N.T., V.S., C.E.S. and J.E.L.), the Angel Fund (R.H.B.Jr.), Project ALS/P2ALS (R.H.B.Jr.), the ALS Therapy Alliance (R.H.B.Jr. and J.E.L.), The Netherlands ALS Foundation (Project MinE; J.H.V. and L.H.v.d.B.), ALS liga Belgium (P.V.D. and W.Ro.), Suna and Inan Kirac Foundation (N.A.B.). Computer resources for this study were provided by the Green High Performance Computing Center at the University of Massachusetts Medical School. L.H.v.d.B. received grants from the Netherlands Organization for Health Research and Development (Vici Scheme; the SOPHIA and STRENGTH projects through the EU Joint Programme – Neurodegenerative Disease Research, JPND). I.P.B. received grant funding from the National Health and Medical Research Council (NHMRC) of Australia (1095215, 1107644). P.C.S. was supported through the auspices of H. Robert Horvitz (Massachusetts Institute of Technology), an Investigator of the Howard Hughes Medical Institute. M.A.v.E. received a grant from the Netherlands Organization for Health Research and Development (Veni scheme) and travel grants from Baxter. This is an EU Joint Programme – Neurodegenerative Disease Research (JPND) project. The project is supported through the following funding organizations under the aegis of JPND (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). C.E.S. and A.A.-C. 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. I.R.C.C.S. Istituto Auxologico Italiano; AriSLA – Fondazione Italiana di Ricerca per la SLA co-financed with support of "5×1000" – Healthcare Research of the Italian Ministry of Health (grants EXOMEFALS 2009 and NOVALS 2012 (N.T., C.T., C.G., V.S. and J.E.L.)), (grant RepeatALS 2013

(S.D.T. and L.C.)), Italian Ministry of Health (grant GR-2011-02347820 – IRisALS (N.T., C.T. and 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 JPND project (STRENGTH). P.V.D. is supported by FWO Vlaanderen and the Belgian ALS liga. In Australia, this work was supported by a Leadership Grant to I.P.B. from MND Australia and an NHMRC fellowship (1092023) to K.L.W. G.A.R. is funded by the Canadian Institute of Health Research (CIHR), Genome-wide exon capture for targeted resequencing in patients with FALS (#208973) the Muscular Dystrophy Association, and Whole exome sequencing in patients with FALS (#153959). We thank JMB Vianney de Jong for collection of clinical data. C.S.L. is recipient of Tim E. Noël fellowship from ALS society of Canada. W.Ro. is supported through the E. von Behring Chair for Neuromuscular and Neurodegenerative Disorders, the Laevers Fund for ALS Research, the ALS Liga België, the fund ‘Een Hart voor ALS’ and the fund ‘Opening the Future’. The research leading to these results has received funding from the European Research Council und the European’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 340429 and from the Geneeskundige Stichting Koningin Elisabeth (G.S.K.E.). M.S. and C.D. were supported by the Deutsche Forschungsgemeinschaft, SE697/4-1, BMBF EnergI and the Deutsche Gesellschaft für Muskelkranke, Project He 2/2. This work was supported in whole or in parts by a grant from the German Federal Ministry of Education and Research (JPND STRENGTH consortium; German network for ALS research MND-NET), the Charcot Foundation for ALS Research, the virtual Helmholtz Institute “RNA-Dysmetabolismus in ALS and FTD” and the DFG-funded Swabian ALS Registry. A.Ch. is funded in part by Italian Ministry of Health (Ricerca Sanitaria Finalizzata 2010, grant RF-2010-2309849, project EXPALS), the European Community’s Health Seventh Framework Programme (FP7/2007-2013 under grant agreements 259867), the Joint Programme – Neurodegenerative Disease Research (Italian Ministry of Education and University) (Sophia, and Strength Projects), A.C. is funded in part by Italian Ministry of Health (Ricerca Sanitaria Finalizzata 2010, grant GR-2010-2320550, project EXTRALS) and Fondazione Vialli e Mauro per la Ricerca sulla SLA onlus (grant #4). FUNDELA – Spanish Foundation to the development of ALS research, ISCIII – Carlos III Institute / Fondo de Investigación Sanitaria of Spain (PI10/00092; PI14/00088), ADELA – ALS Spanish Association. 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 the National Institute on Aging (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 NIH institutes and other foreign governmental and nongovernmental organizations. The Discovery Phase analysis of sequence data is supported through UF1AG047133 to G.D. Schellenberg, L.A. Farrer, M.A. Pericak-Vance, R. Mayeux and J.L. Haines; U01AG049505 to S. Seshadri; U01AG049506 to E. Boerwinkle; U01AG049507 to E. Wijsman; and U01AG049508 to A.M.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 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, 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 US organizations, or foreign governmental or nongovernmental organizations. We thank people with MND, their families and control individuals for their participation in this project.

References

1. Gilissen C, Hoischen A, Brunner HG, Veltman JA. Unlocking Mendelian disease using exome sequencing. Genome Biol. 2011; 12:228. [PubMed: 21920049]
2. Ng SB, et al. Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. Nat Genet. 2010; 42:790–793. [PubMed: 20711175]
3. Smith BN, et al. Exome-wide rare variant analysis identifies *TUBA4A* mutations associated with familial ALS. Neuron. 2014; 84:324–331. [PubMed: 25374358]
4. Jian X, Boerwinkle E, Liu X. *In silico* prediction of splice-altering single nucleotide variants in the human genome. Nucleic Acids Res. 2014; 42:13534–13544. [PubMed: 25416802]
5. Shihab HA, et al. Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. Hum Mutat. 2013; 34:57–65. [PubMed: 23033316]
6. Reid E, et al. A kinesin heavy chain (*KIF5A*) mutation in hereditary spastic paraplegia (SPG10). Am J Hum Genet. 2002; 71:1189–1194. [PubMed: 12355402]
7. Cirulli ET, et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. Science. 2015; 347:1436–1441. [PubMed: 25700176]
8. Brenner D, et al. NEK1 mutations in familial amyotrophic lateral sclerosis. Brain. 2016; 139:e28. [PubMed: 26945885]
9. Renton AE, Chiò A, Traynor BJ. State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci. 2014; 17:17–23. [PubMed: 24369373]
10. Kenna KP, et al. Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing. J Med Genet. 2013; 50:776–783. [PubMed: 23881933]
11. Lattante S, et al. Contribution of major amyotrophic lateral sclerosis genes to the etiology of sporadic disease. Neurology. 2012; 79:66–72. [PubMed: 22722621]
12. Chiò A, et al. Extensive genetics of ALS: a population-based study in Italy. Neurology. 2012; 79:1983–1989. [PubMed: 23100398]
13. Thiel C, et al. NEK1 mutations cause short-rib polydactyly syndrome type majewski. Am J Hum Genet. 2011; 88:106–114. [PubMed: 21211617]
14. Shalom O, Shalva N, Altschuler Y, Motro B. The mammalian Nek1 kinase is involved in primary cilium formation. FEBS Lett. 2008; 582:1465–1470. [PubMed: 18387364]
15. White MC, Quarmby LM. The NIMA-family kinase, Nek1 affects the stability of centrosomes and ciliogenesis. BMC Cell Biol. 2008; 9:29. [PubMed: 18533026]
16. Lee JH, Gleeson JG. The role of primary cilia in neuronal function. Neurobiol Dis. 2010; 38:167–172. [PubMed: 20097287]
17. Lee L. Riding the wave of ependymal cilia: genetic susceptibility to hydrocephalus in primary ciliary dyskinesia. J Neurosci Res. 2013; 91:1117–1132. [PubMed: 23686703]
18. Cohen S, Aizer A, Shav-Tal Y, Yanai A, Motro B. Nek7 kinase accelerates microtubule dynamic instability. Biochim Biophys Acta. 2013; 1833:1104–1113. [PubMed: 23313050]
19. Chang J, Baloh RH, Milbrandt J. The NIMA-family kinase Nek3 regulates microtubule acetylation in neurons. J Cell Sci. 2009; 122:2274–2282. [PubMed: 19509051]
20. Puls I, et al. Mutant dynactin in motor neuron disease. Nat Genet. 2003; 33:455–456. [PubMed: 12627231]
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. 2011; 12:71. [PubMed: 21767396]
22. Chen Y, Craigen WJ, Riley DJ. Nek1 regulates cell death and mitochondrial membrane permeability through phosphorylation of VDAC1. Cell Cycle. 2009; 8:257–267. [PubMed: 19158487]
23. Pelegrini AL, et al. Nek1 silencing slows down DNA repair and blocks DNA damage-induced cell cycle arrest. Mutagenesis. 2010; 25:447–454. [PubMed: 20501547]
24. Sama RR, Ward CL, Bosco DA. Functions of FUS/TLS from DNA repair to stress response: implications for ALS. ASN Neuro. 2014; 6 1759091414544472.

25. Tafuri F, Ronchi D, Magri F, Comi GP, Corti S. SOD1 misplacing and mitochondrial dysfunction in amyotrophic lateral sclerosis pathogenesis. Front Cell Neurosci. 2015; 9:336. [PubMed: 26379505]
26. Madabhushi R, Pan L, Tsai LH. DNA damage and its links to neurodegeneration. Neuron. 2014; 83:266–282. [PubMed: 25033177]
27. Coppedè F, Migliore L. DNA damage in neurodegenerative diseases. Mutat Res. 2015; 776:84–97. [PubMed: 26255941]
28. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium. Identification of genetic factors that modify clinical onset of Huntington's disease. Cell. 2015; 162:516–526. [PubMed: 26232222]
29. Surpili MJ, Delben TM, Kobarg J. Identification of proteins that interact with the central coiled-coil region of the human protein kinase NEK1. Biochemistry. 2003; 42:15369–15376. [PubMed: 14690447]
30. Tryka KA, et al. NCBI's Database of Genotypes and Phenotypes: dbGaP. Nucleic Acids Res. 2014; 42:D975–D979. [PubMed: 24297256]
31. DePristo MA, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011; 43:491–498. [PubMed: 21478889]
32. Fang H, et al. Reducing INDEL calling errors in whole genome and exome sequencing data. Genome Med. 2014; 6:89. [PubMed: 25426171]
33. Raczy C, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. Bioinformatics. 2013; 29:2041–2043. [PubMed: 23736529]
34. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–575. [PubMed: 17701901]
35. Manichaikul A, et al. Robust relationship inference in genome-wide association studies. Bioinformatics. 2010; 26:2867–2873. [PubMed: 20926424]
36. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011; 88:76–82. [PubMed: 21167468]
37. Heinze G, Schemper M. A solution to the problem of separation in logistic regression. Stat Med. 2002; 21:2409–2419. [PubMed: 12210625]
38. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
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). 2012; 6:80–92. [PubMed: 22728672]
40. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013; 7:20. Chapter 7, Unit. [PubMed: 23315928]
41. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4:1073–1081. [PubMed: 19561590]
42. Chun S, Fay JC. Identification of deleterious mutations within three human genomes. Genome Res. 2009; 19:1553–1561. [PubMed: 19602639]
43. Schwarz JM, Rödelsperger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. Nat Methods. 2010; 7:575–576. [PubMed: 20676075]
44. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 39:e118. [PubMed: 21727090]
45. Kircher M, et al. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46:310–315. [PubMed: 24487276]
46. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012; 7:e46688. [PubMed: 23056405]
47. Davydov EV, et al. Identifying a high fraction of the human genome to be under selective constraint using GERP++ PLoS Comput Biol. 2010; 6:e1001025. [PubMed: 21152010]
48. Cooper GM, et al. Distribution and intensity of constraint in mammalian genomic sequence. Genome Res. 2005; 15:901–913. [PubMed: 15965027]

49. Garber M, et al. Identifying novel constrained elements by exploiting biased substitution patterns. Bioinformatics. 2009; 25:i54–i62. [PubMed: 19478016]

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. 2016; 37:235–241. [PubMed: 26555599]

51. Seelow D, Schuelke M, Hildebrandt F, Nürnberg P. HomozygosityMapper—an interactive approach to homozygosity mapping. Nucleic Acids Res. 2009; 37:W593–W599. [PubMed: 19465395]

52. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]

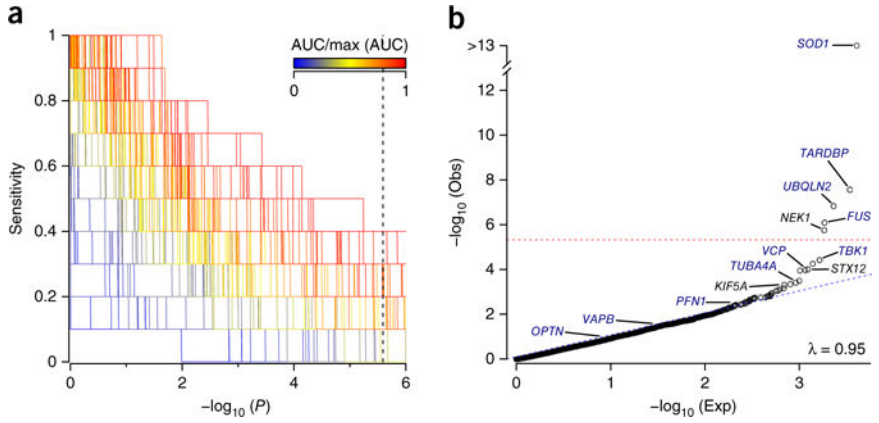

Figure 1.

RVB analysis of FALS exomes. (**a**) RVB analyses of 1,022 index FALS cases and 7,315 controls for 10 known ALS genes, to assess 308 different combinations of MAF 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. λ, observed genomic inflation factor. 'Obs' describes the *P*-value distribution for the observed data. 'Exp' describes the *P*-value distribution under null expectation.

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	
UK	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}

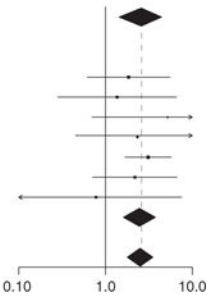

Figure 2.

Replication analysis of *NEK1* p.Arg261His. *NEK1* p.Arg261His genotypes were ascertained for 1,022 FALS samples, 6,172 SALS samples and 11,732 controls. The SALS cohort was divided into seven geographically based case–control strata. Logistic regression was used to conduct tests of allelic association for all subcohorts and was followed by a fixed-effects meta-analysis. In the distribution of OR estimates across study cohorts (right), vertical dotted line denotes OR estimated under meta-analysis. CI, confidence interval.

Table 1

FALS discovery analysis identifies candidate genes

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

RVB analysis results for all genes exhibiting case association at $P < 1 \times 10^{-3}$ in FALS discovery cohort.