Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Landers, John E et al. “Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis.” Proceedings of the National Academy of Sciences 106.22 (2009): 9004-9009. © 2009 National Academy of Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.0812937106">http://dx.doi.org/10.1073/pnas.0812937106</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>United States National Academy of Sciences</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Thu Mar 21 17:12:54 EDT 2019</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/52562">http://hdl.handle.net/1721.1/52562</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis


Amyotrophic lateral sclerosis is a degenerative disorder of motor neurons that typically develops in the 6th decade and is uniformly fatal, usually within 5 years. To identify genetic variants associated with susceptibility and phenotypes in sporadic ALS, we performed a genome-wide SNP analysis in sporadic ALS cases and controls. A total of 288,357 SNPs were screened in a set of 1,821 sporadic ALS cases and 2,258 controls from the U.S. and Europe. Survival analysis was performed using 1,014 deceased sporadic cases. Top results for susceptibility were further screened in an independent sample set of 538 ALS cases and 556 controls. SNP rs1541160 within the KIFAP3 gene (encoding a kinesin-associated protein) yielded a genome-wide significant result (P = 1.84 × 10⁻⁸) that withstood Bonferroni correction for association with survival. Homozygosity for the favorable allele (CC) conferred a 14.0 months survival advantage. Sequence, genotypic and functional analyses revealed that there is linkage disequilibrium between rs1541160 and SNP rs522444 within the KIFAP3 promoter and that the favorable alleles of rs1541160 and rs522444 correlate with reduced KIFAP3 expression. No SNPs were associated with risk of sporadic ALS, site of onset, or age of onset. We have identified a variant within the KIFAP3 gene that is associated with decreased KIFAP3 expression and increased survival in ALS. These findings suggest that genetic factors modify phenotypes in this disease and that cellular motor proteins are determinants of motor neuron viability.

gene-wide association study | single nucleotide polymorphism

Atrophic lateral sclerosis (ALS) is an age-dependent, degenerative disorder of motor neurons (1) that typically develops in the 6th decade and is uniformly fatal, usually within 5 years (2). Approximately 10% of ALS cases are dominantly inherited (3); 20% of these are caused by mutations in the gene encoding copper/zinc superoxide dismutase 1 (SOD1) (4); mutations in the TARDBP gene (5, 6) account for ≈5% of cases. Rare familial cases arise from mutations in genes encoding the vesicle-associated membrane associated protein B (7), also (a RAB5-guanine nu-
variants modify susceptibility, survival, site of onset or age of onset in sporadic ALS, we have undertaken a multicenter genetic analysis of 1,821 sporadic ALS cases (SALS) and 2,258 controls.

**Results**

Genotypes were obtained from 3 sources in the U.S. (Boston, Atlanta, National Institute of Neurological Disorders and Stroke) and 3 in Europe (London, France, Netherlands), using Illumina BeadArrays. Survival information was not available for samples from the NINDS and France. No SNPs generated significant \( P \) values for association with susceptibility, site of onset, and age of onset of disease after Bonferroni correction (288,357 SNPs × 4 phenotypes) (Fig. 1 C–E, Table 1, see also Table S1). In a further attempt to reveal any SNPs that were associated with susceptibility of sporadic ALS, we elected to genotype those SNPs that yielded a \( P < 5.0 \times 10^{-4} \) (153 in total) in a confirmatory (Stage 2) panel consisting of 538 ALS cases and 556 controls. Survival information was not available for most of the samples. Successful genotypes were obtained for 139 (90.8%) of the SNPs; none of the variants yielded a significant \( P \) value after Bonferroni multiple test correction (Table S1). Although our study failed to conform recent reports that susceptibility to sporadic ALS may be mediated by variants in the inositol-triphosphate receptor (ITPR2) (21), DPP6 (22, 23) or a novel, brain-expressed gene (FLJ10986) (24), these discrepancies may reflect differences in methodology or case populations (Table S2).

In contrast to susceptibility, site of onset, and age of onset, SNPs rs1541160 and rs855913 generated significant \( P \) values after Bonferroni correction (288,357 SNPs × 4 phenotypes) for association with disease survival, using linear regression (Fig. 1A and Table 1). For SNP rs1541160, the nominal and Bonferroni-corrected \( P \) values were \( 1.84 \times 10^{-8} \) and 0.021. Within the region of rs1541160, several SNPs (including imputed SNP alleles) yielded a cluster of positive values; 4 of the imputed SNPs were significant after Bonferroni correction (Fig. 1B). SNP rs1541160 maps within intron 8 of the KIFAP3 gene (encoding a kinesin-associated protein) on chromosome 1. For SNP rs855913, the nominal and Bonferroni-corrected \( P \) values were \( 4.02 \times 10^{-8} \) and 0.046. This SNP lies \( \sim \)10 kb upstream of the ZNF746 gene. This gene was not further characterized for 3 reasons. First, a sensitivity analysis of this SNP revealed that it does not replicate within the individual Boston population (\( P = 0.264 \)). Second, in our sensitivity analyses, had we analyzed the U.S. as the Stage 1 population, we would not have identified this variant due to its relatively high \( P \) value (0.0073) and low ranking (2169th). This is in contrast with SNP rs1541160 that emerges as significant in our study, whether considering the aggregate of all cases or each individual population. Finally, for the ZNF746 gene variant in question, the homozygotes for the minor allele are rare (0.7%) so that it is difficult to ascertain the reliability of the results (despite having \( >1,821 \) ALS cases in our screening study).

The genotype frequencies of rs1541160 are 9.9% (CC), 39.7% (CT) and 50.4% (TT); the minor allele frequency is 29.7% (Table S3). The rate of genotyping rs1541160 was 100%. Hardy–Weinberg testing revealed that rs1541160 is in equilibrium (controls \( P = 0.541 \), cases \( P = 0.527 \), all \( P = 0.970 \)). Haplotypes defined by 3 SNPs, rs2750014, rs4656729 and rs12123693, but excluding rs1541160, yielded association with survival comparable to that of rs1541160 (\( P = 1.35 \times 10^{-5} \)), indicating that genotyping artifacts specific to rs1541160 are not generating the association. Further tests confirmed that this association is not biased by population stratification (SI Methods). Pairwise linkage disequilibrium (LD) analysis for \( \sim50 \) SNPs with disease survival, using linear regression (Fig. 1B and Table 1).
SNPs distributed across the locus defined by KIFAP3 and 5 neighboring genes (SCYL3, Clorf156, Clorf112, and Selectins E and L) revealed disequilibrium that spanned ~155 kb from marker rs2750014 to rs1216443 but was centered on rs1541160 within the gene KIFAP3 (Fig. S1).

Our approach to identify variants associated with increased survival was based on a joint analysis of 4 DNA sets. This approach is more powerful than a 2-staged method in which a set of SNPs within an initial population below a cutoff P value is verified within a secondary confirmation population (36). However, because several genome-wide association studies (GWAS) have used 2-stage approaches (21, 22, 24, 30, 32, 33), we have investigated how such an approach would influence our results. We performed a sensitivity analyses (21, 22, 24, 30, 32, 33), we have investigated how such an approach would influence our results. We performed a sensitivity testing (Table 2) revealed that rs1541160 remained the most significant P value associated with both the remaining populations (i.e., simulating a Stage 1 study) and the removed population (i.e., simulating a Stage 2 study). In each case, this sensitivity testing (Table 2) revealed that rs1541160 remained in the top 5 survival-associated SNPs. The largest increase in the P value (to $5.50 \times 10^{-5}$) was observed after removal of the Boston DNA set, which contains the most samples ($n = 558$). Furthermore, the P values for rs1541160 from each individual population (the simulated Stage 2 study) also yielded significant P values, with the exception of that from Atlanta ($P = 0.079$), which contains the lowest number of samples ($n = 90$). Interestingly, the median survival increased at a minimum of 20.51% (Netherlands) and a maximum of 78.54% (London) in individuals harboring a CC genotype for rs1541160 as compared with a TT genotype (Table 2). We also performed the sensitivity analysis by grouping the population into U.S. (Atlanta and Boston) and Europe (London and Netherlands). By this approach, both the U.S. and Europe yielded a high ranking for rs1541160 if used as a Stage 1 population (20th and 31st, respectively). Furthermore, both the U.S. and Europe yielded significant individual P values ($5.55 \times 10^{-5}$ and $7.70 \times 10^{-5}$, respectively) (Table 2). These results also confirm that the observed association is not due to population stratification, which would not be expected to yield a significant $P$ value for each individual population.

The absolute median survival for the CC, CT, and TT genotypes were 3.96, 2.84 and 2.67, respectively. The absolute mean survival for the CC, CT, and TT genotypes were 4.60, 3.40, and 3.07, respectively. As assessed by linear regression analysis, the mean and median survival increments for the CC genotype were 14.0 and 14.9 months, respectively, compared with the TT genotype, based on the analysis with SNP rs1541160 alone. With genotypic-based survival curve analysis using the Peto-Prentice generalized Wilcoxon test and deceased ALS cases (Fig. 2A), the P value for the rs1541160 SNP is $3.87 \times 10^{-5}$ ($n = 1,014$). A censored analysis considering all of the cases (Fig. 2B) yielded a $P$ value for rs1541160 of $1.82 \times 10^{-5}$ ($n = 1,521$).

Sequence analysis of the KIFAP3 coding region and exon/intron boundaries of 8 individuals homozygous for the CC and 4 for the TT rs1541160 genotype (12 individuals) did not reveal variants in strong LD with rs1541161, suggesting that the KIFAP3-mediated increase in survival is not due to an alteration in its protein sequence. To determine whether the expression of KIFAP3 is modified by the genotype of rs1541160, we performed real-time PCR on lymphoblastoid cell lines harboring either a CC ($n = 38$) or TT ($n = 40$) genotype for rs1541160. KIFAP3 expression in the CC genotypes was 31.9% less than that in the TT genotypes (Fig. 3A) ($P = 0.0084$, Wilcoxon 2-sample test). A comparison of

| Table 2. Sensitivity analysis of rs1541160 within four populations. Median Survival is represented in years |
|---|---|---|---|---|---|---|---|
| Stage 2 population | Stage 1 P | Stage 1 rank | Stage 2 P | Stage 1 sample size | Stage 2 sample size | Stage 2 median surv. CC | Stage 2 median Surv. TT |
| Stage 2 survival increase (CC vs. TT) |
| Atlanta | 2.89E-07 | 1st | 0.079 | 924 | 90 | 4.00 | 2.93 |
| Boston | 5.50E-06 | 5th | 8.86E-04 | 616 | 398 | 4.27 | 2.61 |
| London | 4.00E-06 | 5th | 1.42E-03 | 804 | 210 | 4.66 | 78.54 |
| Netherlands | 2.60E-07 | 1st | 0.022 | 698 | 316 | 2.82 | 2.24 |
| United States | 7.70E-05 | 31st | 5.55E-05 | 526 | 488 | 4.23 | 2.85 |
| Europe | 5.55E-05 | 20th | 7.70E-05 | 488 | 526 | 3.39 | 40.08 |
| Total | $1.84 \times 10^{-5}$ | | | 1,014 | | 3.96 | 2.67 |

9006 | www.pnas.org/cgi/doi/10.1073/pnas.0812937106

Landers et al.
occipital lobe brain samples homozygous for either the C (n = 9) or T (n = 17) alleles again revealed a decrease in expression of KIFAP3 (41.1%) in the CC as compared with TT samples (Fig. S2B) (P = 0.025, Wilcoxon 2-sample test). A second real-time PCR probe for KIFAP3 confirmed the findings for both lymphoblast and brain samples. Western blotting of the brain samples revealed a variant (C/G, previously identified as rs522444) located at 25 bp relative to the transcription start site (Fig. S2A) uncensored data; (B) censored data. In B, small vertical marks superimposed on the survival curve show censored points at which individuals were lost to further assessment.

Given that rs1541160 is located in the eighth intron of KIFAP3, we decided to examine SNPs within the promoter region which may influence gene regulation to determine whether any were in linkage disequilibrium with rs1541160. Sequencing of ~1.6 kb of this promoter within 8 individuals homozygous for the CC and 4 for the TT rs1541160 genotype revealed a variant (C/G, previously identified as rs522444) located at ~25 bp relative to the transcription start site (Fig. 3C). This variant is in complete linkage disequilibrium with rs1541160 for all 24 chromosomes (r² = 1.00). Genotypes derived from the HapMap project for rs1541160 and rs522444 further documented that these 2 SNPs are in complete LD (r² = 1.00) in Caucasian and Sub-Saharan African populations (evident through analysis of 116 and 118 chromosomes, respectively). We further confirmed this high level of LD by genotyping an additional 1,017 individuals (r² = 0.99).

Analysis of the KIFAP3 promoter using transcription element search system (TESS) (37) revealed that the KIFAP3 gene lacks a TATA box and that rs522444 lies within a putative Sp1 binding site (log-likelihood score = 12) (Fig. 3C). Moreover, the C allele of rs522444, which is in linkage disequilibrium with the lower expressing C allele of rs1541160, creates the putative Sp1 binding site, whereas the G allele at rs522444 destroys this site. Comparison with the promoter region in other primates demonstrates that the G allele is evolutionarily conserved suggesting this is the ancestral allele and that the KIFAP3 gene is not normally regulated by a Sp1 binding site. Because Sp1 family members binding to cognate Sp1 brain tissue harboring either the CC or TT genotype and subcloned upstream of the firefly luciferase gene. The schematic (not drawn to scale) represents the resultant constructs, which differ only at a single base pair located at rs522444. (E) The resultant constructs were transfected into SK-N-AS cells and relative luciferase activity was measured. The error bars represent the 95% C.I. A promoterless vector yielded <1% relative activity. The construct containing the G allele displays higher luciferase activity relative to the C allele. *, P < 0.05; **, P < 0.01.

Discussion

In a set of 4,079 DNA samples from sporadic ALS cases and controls we have completed an unbiased analysis of ~288,000 SNPs distributed across the genome and identified a SNP (rs1541160) within the KIFAP3 gene that is associated with reduced KIFAP3 expression and longer survival. Other SNPs in the region of
rs1541160 trended toward association (Fig. 1B); the 4th and 10th highest SNPs in the survival analysis were also within the KIFAP3 gene (Table S1). The failure to detect other significant SNP genotypes is subject to multiple interpretations. Perhaps most importantly, this suggests that there is not a single, readily detectable genetic variant that exerts a preponderant influence on either the risk of developing sporadic ALS or ALS phenotypes other than survival. In the present study, the absence of strongly associated SNPs other than rs1541160 may reflect other factors including inherent heterogeneity in the populations studied, locus and allelic heterogeneity, the inability of our present study design to detect underlying epistatic interactions of multiple gene variants, the effect of a microdeletions or insertions or inadequacies in the power of our study to detect genes of small effect. (For susceptibility studies, underling epistatic interactions of multiple gene variants, the effect heterogeneity, the inability of our present study design to detect important, this suggests that there is not a single, readily detectable gene (Table S1). The failure to detect other significant SNP

That homozygosity for SNP rs1541160 confers survival variation of ~14 months is of clinical importance in a disorder with a mean survival of only ~3–5 years. Why attenuation of KIFAP3 expression should be protective in ALS is unclear. With the kinesin motor proteins KIF3A and KIF3B, KIFAP3 forms a trimetric motor complex, KIF3, that mediates binding between the motor proteins and their cargoes, serving multiple functions such as chromosomal cytokinesis and anterograde transport (39, 40). Presumably, reduced levels of KIFAP3 modulate survival by favorably affecting both the stoichiometry of KIFAP3 and the KIF3 complex and one or more transport functions, such that the CC genotype of rs522444 is beneficial. Heightened expression of KIFAP3 is reportedly an early marker of disease in transgenic mutant SOD1 mice (41), suggesting that levels of KIFAP3 reflect adverse events within motor neurons. In human sporadic ALS, expression of 2 other kinesin-related proteins (KIF3Ab and KIF1Bb) is reportedly reduced (42). A recent report documents that KIFAP3 (also described as KAP3) binds mutant SOD1 protein, slowing axonal transport of choline acetyl transferase in motor neurons; moreover, KIFAP3/KAP3 colocalizes with mutant SOD1 in human motor neuron at autopsy (43). It is conceivable that diminished expression of KIFAP3/KAP3 has a beneficial impact on the SOD1-motor protein interaction in ALS. More generally, mutations in motor proteins are implicated in multiple motor neuron degenerative disorders in both humans (11, 44) and mice (45).

It is encouraging that investigations employing complex genetics may provide fresh insight into sporadic ALS (21–24, 46). Few genetic factors that modify ALS survival are reported (19, 20, 47); none were identified in the previous ALS genome analyses (21–24, 46). The identification of KIFAP3 as a determinant of progression rate of sporadic ALS is therefore promising; insights into this pathway may provide new targets for therapies to slow this devastating disease, for example, by reducing levels of KIFAP3 expression or modifying its interactions with 1 or more protein binding partners.

Materials and Methods

Genotypes were obtained from 3 sources in the U.S. (917 ALS, 912 controls) and 3 in Europe (904 ALS, 1,346 controls) (Table S5). To maximize the power of this study, we combined these into a single set of cases and controls (36). Duration information was obtained from 4 of the sites. A set of 307,776 SNPs common to all sites was used for this analysis. Multiple quality control measures were applied to the set of DNAs and SNPs. 10,360 SNPs were eliminated because they were not in Hardy–Weinberg equilibrium (p < 10−6 in controls) or demonstrated call rates <0.95 or minor allele frequencies <0.001. An additional 9,059 SNPs were eliminated by tests for divergence of cases and control call rates and for nonrandom missing genotype data (to determine whether genotypes are missing with respect to the true genotype as defined by the observed genotypes of nearby SNPs). Samples were excluded if their call rates were <95%; if genotypes revealed duplicate samples, relatedness (proportion of genome IBD > 0.2), excess homozygosity or heterozygosity (inbreeding coefficient >0.05, excess heterozygosity >0.025); or if they represented duplicate samples, genotypically-assessed gender. The sample set was additionally subjected to stratification analysis; based on the distribution of pairwise genome-wide identity-by-state distances, we applied complete linkage hierarchical cluster analysis and classical multidimensional scaling. As a result, 72 outlier samples, defined as 3 standard deviations from the group mean, were eliminated, leaving the cases and controls as in Table S5. For the final analysis, 288,357 SNPs were evaluated (Table S6). After applying quality control metrics to the full set of DNAs and SNPs (SI Methods), 288,357 SNPs remained that were evaluated in 4,079 DNA samples, yielding 1,176,208,203 genotypes (Table S6). Genotypic and phenotypic data are available through the dbGaP database at the National Center for Biotechnology Information web site (www.ncbi.nlm.nih.gov/). Genotypes were used for the analysis of 4 SALS phenotypes: susceptibility, site of onset, age of onset and survival of disease. Multiple analyses failed to detect biases introduced by stratification of our case-control cohorts (see SI Methods).

ACKNOWLEDGMENTS. We dedicate this article to the memories of Frieda R. Abraham, Donna Roberts, and Jennifer Estess. We thank Daniel Mirel (Broad Institute) for all of his assistance, all of the clinicians who assisted in recruiting cases and controls, particularly to clients with ALS and their families, and Jerry Xu and Dennis Gurgul of the Enterprise Research IS group at Partners Healthcare for support and use of the Hewlett-Packard computing facilities. The Boston-based investigations were supported by the ALS Therapy Alliance, Project ALS, the Department of Health and the Pierre L. de Lattre Fund, the ALS Research Foundation, the Alf-Athena ALS Research Foundation (R.H.B., J.L.E., D.M.M.-Y., L.R., T.T.K.), the ALS Family Charitable Foundation, the American Academy of Neurology Foundation/ALS Association (A.-M.W.), the National Institute of Neurological Disorders and Stroke (R.H.B., T.T.K.), the Harvard Business School Class of 2007, the Yammer family, Avichai Kremer, Guy Yamen, Nate Boaz, and Professor Robert S. Kaplan. London-based investigations were supported by the Motor Neurone Disease Association of Great Britain and Northern Ireland (A.A.-C.,), the Medical Research Council (A.A.-C.,), the Wellcome Trust (P.N.L. and C.E.S.), and Psychiatry Research Trust (Tim Perkins Fund) (P.N.L., A.A.-C., and C.E.S.). The investigations in France were supported by Assistance Publique, Hopitaux de Paris, Institut National de la Santé et de la Recherche Médicale, GIS-Institut des Maladies Rares and the Ministère de l’Enseignement Supérieur et de la Recherche. V.S. was partially supported by a donation of the Peviani Family. We thank DNA banks maintained by the Coriell institute (with support from the National Institute of Neurological Disorders and Stroke) and Genethon. This study used data from the SNP Database at the National Institute of Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds). Investigations in The Netherlands were supported by The Netherlands Organisation for Scientific Research, the “Prinses Beatrix Fonds,” the Kersten Foundation, the Adessium Foundation, J. R. van Dijk and VSB Funds (L.H.v.d.B.). This work was supported by Howard Hughes Medical Institute (P.S. and H.R.H.), the Packard Center for ALS Research (J.D.G. and M.P.), The Muscular Dystrophy Association (U.S.) (S.C. and O.H.), and National Center for Research Resources Grant U54 RR020278–01 (to The Broad Institute Center for Genotyping and Analysis). H.R.H. is Howard Hughes Medical Institute Investigator.


