Supporting Information

Landers et al. 10.1073/pnas.0812937106

SI Materials and Methods

Sample Selection and Genotyping. Genotyping was performed with Illumina BeadArrays at the Broad Institute in Boston (Boston-Atlanta-London DNA set, HumanHap300 v.1.0), the National Institutes of Health (Drs. A. Singleton and J. Hardy, NINDS/Coriell DNA set; http://ccr.coriell.org/ninds) (1, 2), the Centre National de Genotypage at Evry (French set, HumanHap300 v.1.0 and HumanHap300-Duo v.2) and at Rudolf Magnus Institute of Neuroscience, Utrecht (Netherlands set, HumanHap300 v.1.1). All cases were Caucasian and of European-American descent and fulfilled the El Escorial criteria for probable or definite ALS (3). The SALS DNA set encompassed 1,090 males and 731 females (Table S5) with mean ages of 55.6 ± 12.2 and 59.2 ± 12.5 years (Table S5). In the DNA sets from Boston, NINDS, Atlanta, London and the Netherlands, case and control samples were matched for gender and age distribution; because there were no controls from Atlanta, the controls in the Boston set were selected so that the aggregated Boston-Atlanta case-control sets matched as fully as possible for age and sex. However, because the French controls were nearly all males, the ratio of males to females in the combined set is smaller for the ALS cases (1.49) than the controls (2.30). Distributions of sites of onset were comparable in the different ALS cohorts (Table S5). The NINDS collection is a publicly accessible resource containing whole genome genotypes on sporadic ALS samples and controls. This dataset is maintained at the Coriell Institute (http://ccr.coriell.org/ninds). To optimize statistical power, all NINDS case and control genotypic data available at the time of the study were acquired and analyzed as a single set. The DNA sets were screened to exclude acquisition of samples present in more than one collection. As a result, the NINDS case and control samples do not have an equal sex ratio. Patients with a family history of ALS were excluded. As a test of the utility of the family history, *SOD1* mutation analysis (via exon sequencing) was performed on 184 cases in the Boston SALS DNA set; 4/184 cases had *SOD1* gene mutations. Because only 4 mutations were identified (2.2%), we did not perform mutational screening on the remaining samples; all cases with known *SOD1* mutations or a known family history of ALS were excluded.

Statistical Analyses. All genotypic analyses were conducted using the computer package PLINK (4) (http://pngu.mgh.harvard.edu/purcell/plink). To augment the power of this study, we elected to consider all of the ALS data as a single case series, and the combined controls as a single control DNA set; on theoretical grounds, this approach optimizes the likelihood of finding subtle associations for a given case-control sample size (36). Susceptibility was defined simply as having definite or probable ALS, as defined by El Escorial criteria (3). Age of onset was defined by the onset of first symptoms. The survival endpoint was death or time of initiation of all forms of invasive ventilatory support. Sites of onset for the cases were registered as limb vs. bulbar; the statistical analysis compared SNP allele frequencies in the population of bulbar vs. limb onset cases.

To control for residual stratification in the single SNP screening, individuals' scores on the first 2 dimensions from the multidimensional scaling analysis were entered as covariates (Fig. S3), in either a logistic model for binary traits (susceptibility and site of onset) or linear regression model for quantitative traits (age of onset and survival). Site of collection (dummy-variable coded) was also entered as a covariate. Bonferroni correction for multiple testing was performed using the total number of SNPs tested (288,357) times the 4 phenotypes. Complete GWAS *P* values are available for download.

Genotype imputation was performed within a 10Mb region surrounding the *KIFAP3* gene. Genotypes were imputed at SNPs that were in HapMap but not directly genotyped using the program IMPUTE v0.3.0 (www.stats.ox.ac.uk/~marchini/software/gwas/gwas.html). As recommended by the IMPUTE website, an effective population size of 11,418 was used for imputation. All other parameters were left at their defaults. Imputation was based on HapMap Phase II rel#21 data files (NCBI build 35). Genotypes with probabilities below the default call threshold of 0.9 were treated as missing. Imputed SNPs with a minor allele frequency <0.001 were excluded. Genotypic analyses were conducted using PLINK.

For the survival analyses (including both the linear regression analysis and the Wilcoxon-based survival analysis), the French and the NINDS cases were not used, the former because they included only those cases that survived <5 years, potentially biasing the set against detection of survival-influencing alleles, and the latter because survival data were not available. Duration data were available on 307 living and 1,014 deceased SALS cases (Table S5). Our initial screen for association with survival used linear regression, using only deceased individuals. We secondarily performed a Peto-Prentice generalized Wilcoxon test of genotypes and disease duration (using both living and deceased patients). This test was chosen due to the higher frequency of early deaths. As shown in Table S5, the mean and median survival periods of disease were 3.35 and 2.79 years. Interestingly, in the series from the Netherlands, the mean and median survival periods were 2.60 and 2.34 years. The majority of ALS cases were deceased (Table S5). The Peto-Prentice generalized Wilcoxon Testing and Wilcoxon Two Sample Testing were performed using the software package SAS 9.1 (SAS Institute).

Stratification Analyses. Because our study entails analysis of 6 different case-control cohorts from 4 countries, we performed multiple tests to detect biases introduced by stratification. These tests and related considerations document that our analysis is not confounded by stratification. (*i*) Our initial SNP analysis controlled for residual stratification by using each individual's scores from the first 2 dimensions of a multidimensional scaling analysis of IBS distances as covariates. (*ii*) Testing based on the χ^2 sum statistic for 1,000 randomly chosen, unlinked SNPs demonstrated that the populations were genetically similar ($P = 0.36$). (*iii*) A Q-Q plot for the genome-wide distribution of negative log *P* values for the test of association with survival shows no systematic bias, with greater-than-expected levels of association only at the extreme tail of the distribution due to SNPs located within the *KIFAP3* and *ZNF746* gene regions (Fig. S4). (*iv*) The association analysis of the 3 other phenotypes does not show any evidence for false positives induced by stratification.

For studies of SNP rs1541160, additional lines of evidence further excluded stratification effects. (*i*) The *P* values for rs1541160 from each individual site were computed and found to yield significant *P* values, with the exception of that from Atlanta ($P = 0.079$), with the lowest number of samples ($n = 90$) (Table 2). If our overall results reflected population stratification, one would not expect each population to display an individually significant *P* value. (*ii*) We performed a sensitivity analysis, dropping each of the 4 populations in turn from the study. If the overall *P* value reflects the impact of a single population, then its elimination should yield a nonsignificant *P* value. However,

our P value remained significant despite removal of any one of the 4 populations (Table 2). The largest increase in P value (to 5.50×10^{-6}) was observed after removal of the Boston DNA set, with the most samples ($n = 398$). (*iii*) We tested whether rs1541160 is correlated with either dimension of stratification plots (Fig. S3). That is, we treated the X and Y values individually, as a phenotypic trait and tested for association with rs1541160. Neither showed a significant value ($P = 0.676$ and 0.150, respectively) indicating there is no evidence that the SNP varies by the population structure as determined by the stratification plots. The following relevant points are also noted. (*iv*) Hardy-Weinberg testing revealed that rs1541160 is in equilibrium in each population used. (*v*) Each population demonstrated a similar allele frequency for rs1541160 confirming that the observed association is not due an idiosyncrasy of this parameter in a single population (Table S2). (*vi*) Adjusting our analysis for age of onset, site of onset and gender also did not substantially alter the association ($P = 9.1 \times 10^{-8}$).

Expression Analysis, Luciferase Assays, and Western Blot Analysis. Total RNA was reverse-transcribed using SuperScript III First-Strand Synthesis kit (Invitrogen) according to the manufacturer's protocol. All RNA samples were confirmed to be intact by gel electrophoresis. Briefly, cDNA synthesis was performed in a 20-μL reaction containing 1× RT Reaction mix, 0.5 μg of total RNA and 2 μL of SuperScript III. The reaction was incubated at 25 °C for 10 min and heated to 50 °C for 30 min followed by 85 °C for 5 min. *E.coli* RNase H (2 units) was added and the reaction was incubated at 37 °C for 20 min. Probes Hs00946074 (shown) and Hs00183973 (Applied Biosystems) were used for the quantization of *KIFAP3* and a probe for beta-2-microglobulin was used as an endogenous internal control (Applied Biosystems, cat no. 4326319E). Quadruplicate reactions were performed in a 20-μL volume containing 1× Platinum PCR Supermix (Invitrogen), 1× target probe mix and 1× endogenous probe mix. Real-time PCR was performed under the following cycling conditions: 50 °C for 2 min; 95 °C for 2 min.; 40 cycles of 95 °C for 15 sec, 60 °C for 1 min. Reactions were initially normalized using a ROX passive reference and then relative quantization was established to a calibrator sample after normalizing to the endogenous internal control, using the software application SDS 2.2 (Applied Biosystems).

The luciferase constructs were created by amplifying the 633 bp of *KIFAP3* promoter and 5′ UTR, using Platinum Pfx DNA polymerase (Invitrogen) with primers PROM-F (CGCGCGGTACCggttgggggaggtgtattct) and PROM-R (GCGCCTCGAGTGGAGAGGATGGGGTATCTTG). The template for the PCR was genomic DNA derived from individuals homozygous for either the G or C allele of rs522444. The amplified products were digested with the restriction enzymes KpnI and XhoI then subcloned into the firefly luciferase reporter vector pGL3-Basic (Promega). The insert of the resultant constructs was sequenced to ensure no other differences were present. Transfections were performed with Lipofectamine2000 on 80–90% confluent SKN-AS cells in 96-well plates, using 100 ng of each construct, according to the manufacturer's recommendations. The cells were also cotransfected with 20 ng of pRL-SV40 encoding *Renilla* luciferase, to act as an internal control reporter. After 24 h, firefly and *Renilla* luciferase activity was measured in a Victor^3 V 1420 Multilabel Counter, using the Dual-Luciferase Reporter Assay System (Promega). After background subtraction, firefly luciferase activity was normalized using *Renilla* luciferase activity and the relative activity was determined. The results shown are from 4 experiments with at least triplicate wells used for each experiment. The firefly luciferase reporter vector pGL3-Basic yielded <1% activity relative to the *KIFAP3* promoter containing constructs.

Brain lysates were prepared by adding 500 μL of RIPA buffer with protease inhibitor (Protease Inhibitor Mixture Tablets, Roche Applied Science) to ≈50 mg of tissue followed by dounce homogenization. The lysate was then sonicated by 20 × 1 second pulses and then kept on ice for 30 min. After centrifugation, the supernatant was stored in aliquots at −80 °C. The concentration of each sample was determined by a BCA assay. Fifty micrograms of lysate were separated by Tris-glycine SDS/PAGE, and blotted onto PVDF membrane. The membranes were hybridized with a monoclonal antibody direct against KIFAP3 (cat. no. sc-55598, Santa Cruz Biotechnology) at a 1:100 dilution and a rabbit polyclonal antibody directed against Actin (cat. no. 2066, Sigma Aldrich) at a 1:2,000 dilution. Detection was performed using the SuperSignal West Femto Maximum Sensitivity Substrate (Pierce Biotechnology) in conjunction with a CCD camera apparatus. Quantitation of band intensity was performed using the ImageJ software.

1. Schymick JC, et al. (2007) Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: First stage analysis and public release of data. *Lancet Neurol* 6:322–328.
2. Simon-Sanchez J, et al. (2007) Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet* 16:1–14.
3. Brooks B (1994) El Escorial World Federation of Neurology Criteria for the Diagnosis of Amyotrophic Lateral Sclerosis. *J Neurol Sci* 124:96–107.
4. Purcell S, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575.

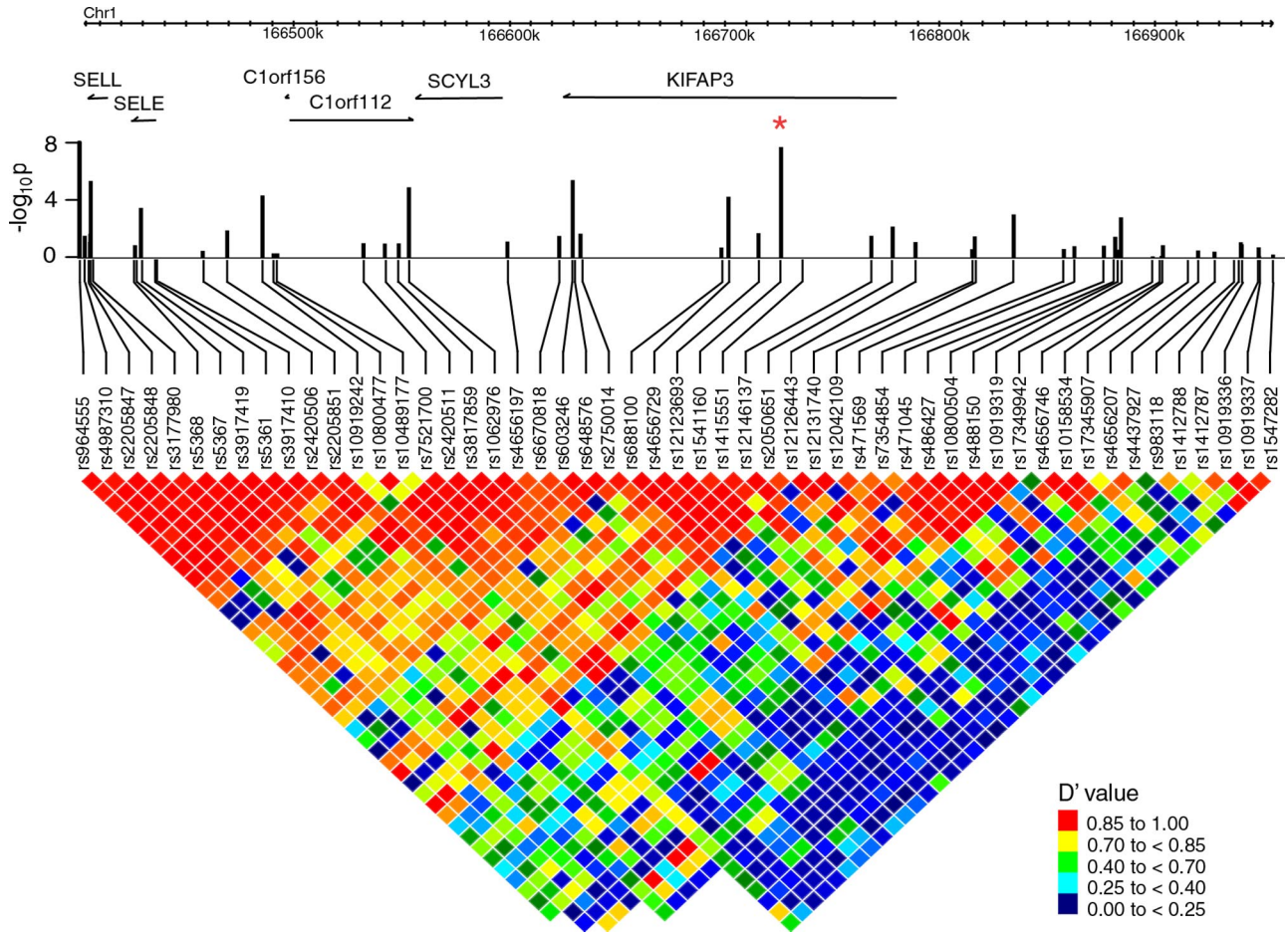

Fig. S1. Linkage disequilibrium plot for SNP rs1541160/*KIFAP3* locus. Pairwise linkage disequilibrium values (D′) were calculated for SNPs spanning the *KIFAP3* region. The color key for D′ values is shown. The $-\log_{10}(P)$ values established by linear regression analysis for duration are shown above each SNP. An asterisk indicates the rs1541160 SNP. The location of additional genes in the region is shown.

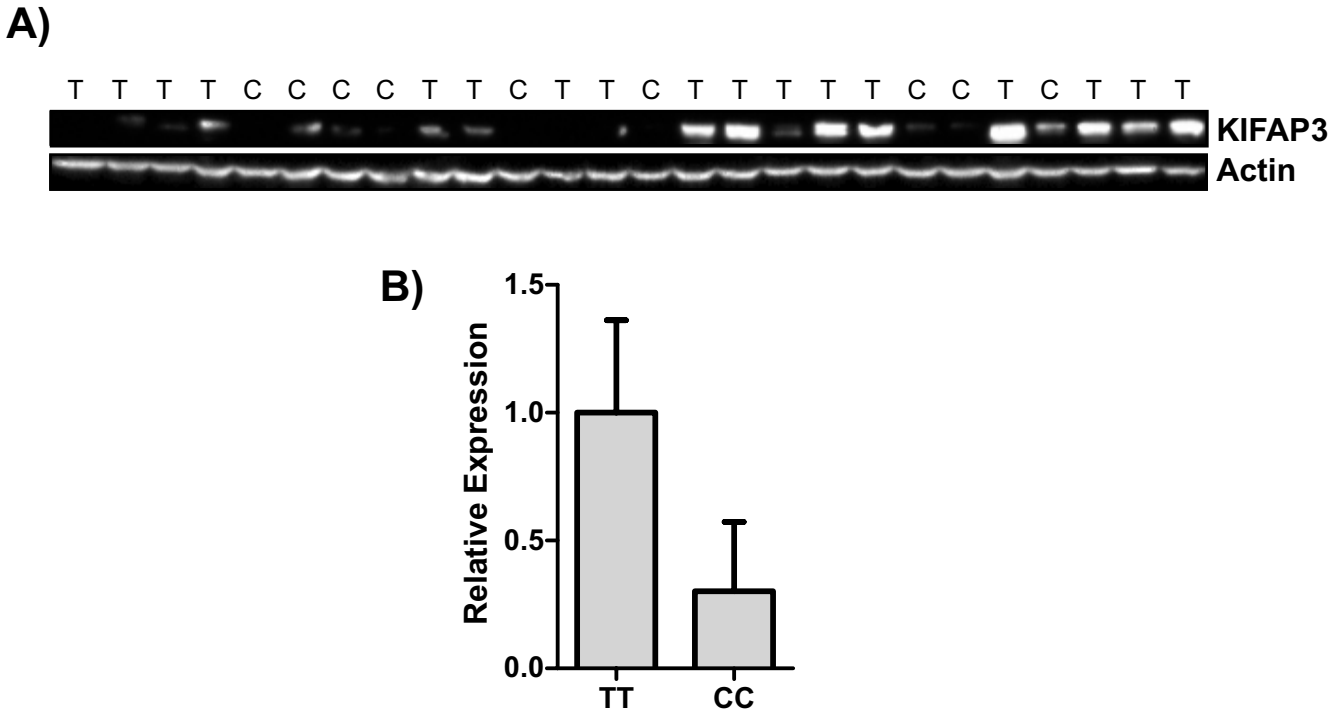

Fig. S2. Association of rs1541160/rs522444 with KIFAP3 protein expression. Total protein was isolated from occipital cortex brain tissue harboring either a CC or TT genotypes for rs1541160. (*A*) Relative expression of KIFAP3 was determined by Western blot analysis followed by quantitation, using ImageJ software. As shown in *B*, individuals harboring the CC genotype display decreased expression (69.8%) compared with individuals with the TT genotypes. Error bars represent the 95% C.I.

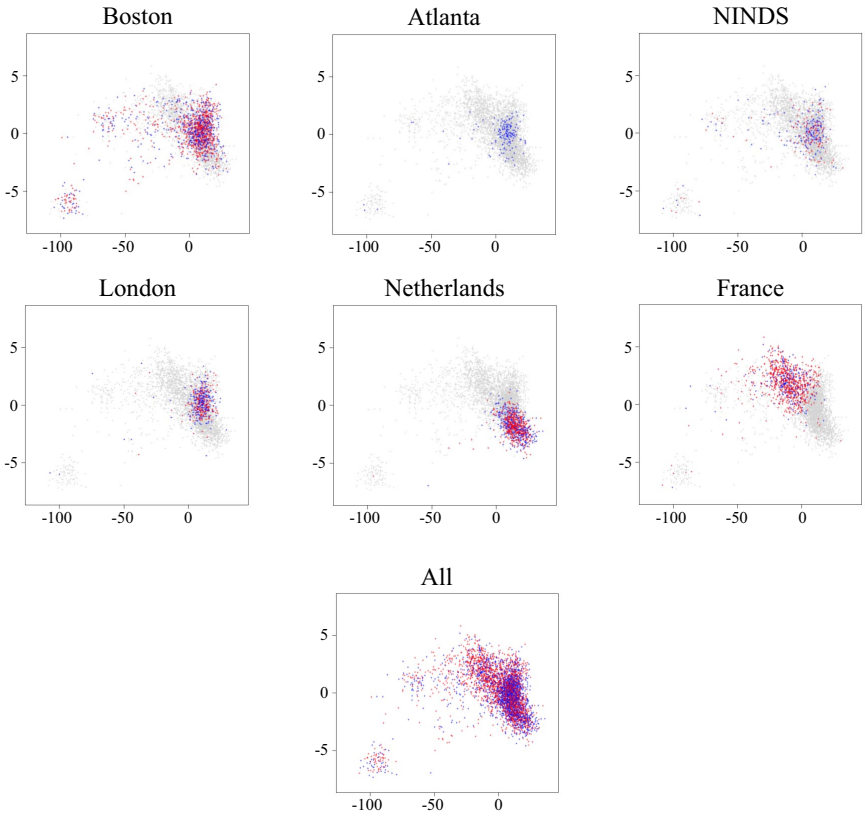

Fig. S3. Stratification plots for each DNA set. The sample sets was subjected to stratification analysis using the computer package PLINK. Pairwise genome-wide identity-by-state distances were calculated and complete linkage hierarchical cluster analysis and classical multidimensional scaling were applied. Outlier samples, defined as >3 standard deviations from the group mean, were eliminated from further analysis. The remaining samples are plotted using the first 2 dimensions from the multidimensional scaling analysis. Blue and red points show case and control samples, respectively, for the DNA set indicated; gray points indicate all samples.

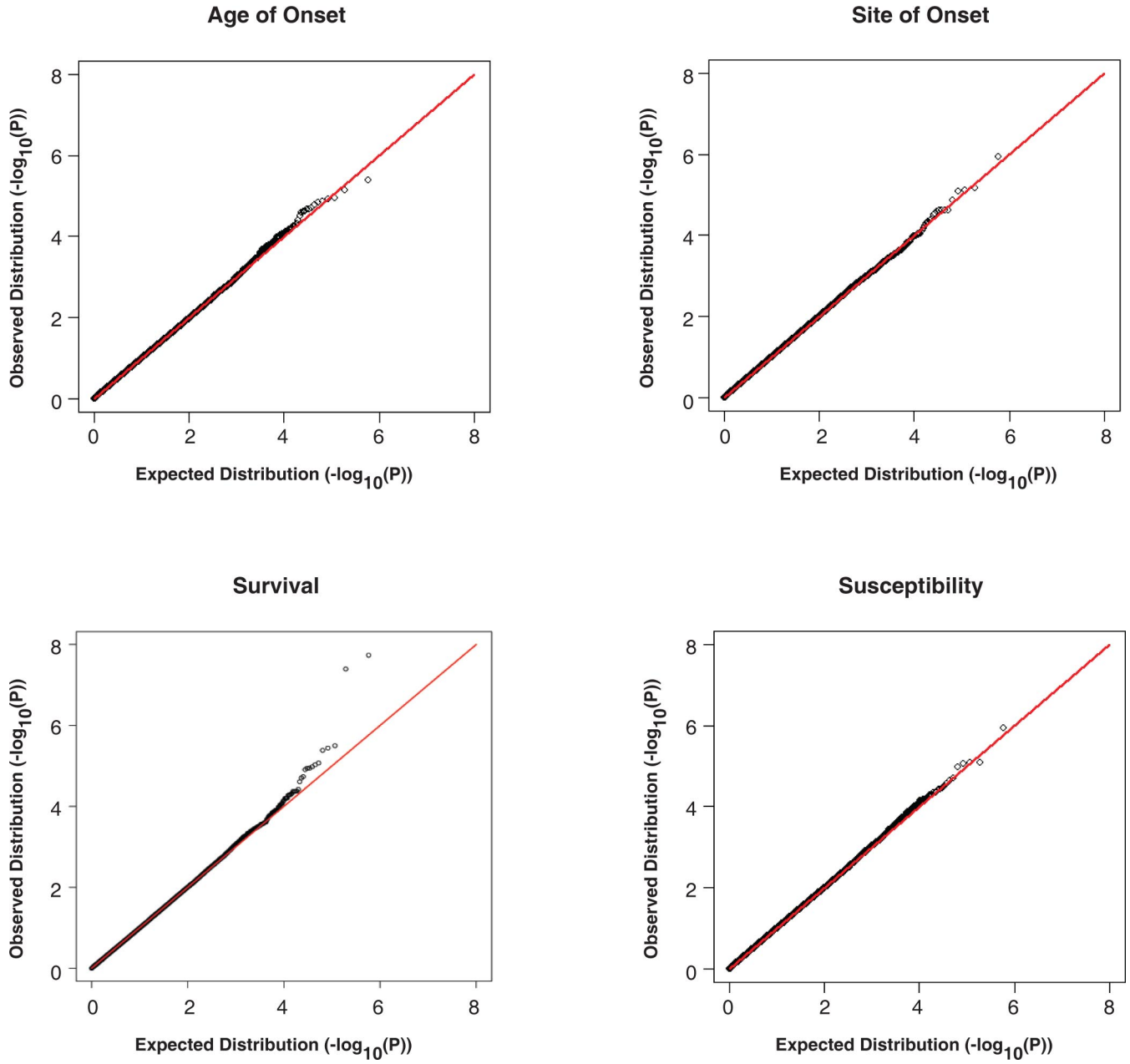

Fig. S4. Observed and expected distributions of *P* values in genome-wide scans for association with age of onset, site of onset, survival, and susceptibility in sporadic ALS. The observed *P* values for the 4 phenotypes indicated were plotted against their expected distribution. The red line indicates the plot expected under the null hypothesis.

Other Supporting Information Files

Table S1
Table S2
Table S3
Table S4
Table S5
Table S6