

A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis

Adriano Chiò^{1,†}, Jennifer C. Schymick^{2,3,†}, Gabriella Restagno^{4,†}, Sonja W. Scholz^{5,6}, Federica Lombardo⁴, Shiao-Lin Lai^{5,7}, Gabriele Mora⁸, Hon-Chung Fung^{2,7}, Angela Britton⁵, Sampath Arepalli⁵, J. Raphael Gibbs^{6,9}, Michael Nalls⁵, Stephen Berger², Lydia Coulter Kwee^{10,11}, Eugene Z. Oddone^{11,12}, Jinhui Ding⁹, Cynthia Crews², Ian Rafferty², Nicole Washecka², Dena Hernandez^{5,6}, Luigi Ferrucci¹³, Stefania Bandinelli¹⁴, Jack Guralnik¹⁵, Fabio Macciardi¹⁶, Federica Torri¹⁶, Sara Lupoli¹⁷, Stephen J. Chanock¹⁸, Gilles Thomas¹⁸, David J. Hunter^{18,19}, Christian Gieger^{20,21}, H. Erich Wichmann^{20,21}, Andrea Calvo¹, Roberto Mutani¹, Stefania Battistini²², Fabio Giannini²², Claudia Caponnetto²³, Giovanni Luigi Mancardi²³, Vincenzo La Bella²⁴, Francesca Valentino²⁴, Maria Rosaria Monsurrò²⁵, Gioacchino Tedeschi²⁵, Kalliopi Marinou⁸, Mario Sabatelli²⁶, Amelia Conte²⁶, Jessica Mandrioli²⁷, Patrizia Sola²⁷, Fabrizio Salvi²⁸, Iliara Bartolomei²⁸, Gabriele Siciliano²⁹, Cecilia Carlesi²⁹, Richard W. Orrell³⁰, Kevin Talbot³, Zachary Simmons³¹, James Connor³², Erik P. Piro³³, Travis Dunkley³⁴, Dietrich A. Stephan³⁴, Dalia Kasperaviciute³⁵, Elizabeth M. Fisher³⁵, Sibylle Jabonka³⁶, Michael Sendtner³⁶, Marcus Beck³⁶, Lucie Bruijn³⁷, Jeffrey Rothstein³⁸, Silke Schmidt^{10,11}, Andrew Singleton⁵, John Hardy^{2,6} and Bryan J. Traynor^{2,38,*}

¹Department of Neuroscience, University of Turin, Turin, Italy, ²Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD, USA, ³Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK, ⁴Molecular Genetics Unit, Department of Clinical Pathology, A.S.O. O.I.R.M.-S. Anna, Turin, Italy, ⁵Molecular Genetics Unit, Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD, USA, ⁶Department of Molecular Neuroscience and Reta Lila Weston Institute of Neurological Studies, Institute of Neurology, Queen Square, London, UK, ⁷Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Taiwan, Republic of China, ⁸Salvatore Maugeri Foundation, Lissone, Italy, ⁹Computational Biology Core, Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD, USA, ¹⁰Center for Human Genetics, Duke University Medical Center, Durham, NC, USA, ¹¹Department of Medicine, Duke University Medical Center, Durham, NC, USA, ¹²Epidemiology Research and Information Center, Durham VAMC, Durham, NC, USA, ¹³Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, MD, USA, ¹⁴Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy, ¹⁵Laboratory of Epidemiology, Demography and Biometry, National Institute on Aging, Bethesda, MD, USA, ¹⁶Department of Science and Biomedical Technology, University of Milan, Milan, Italy, ¹⁷INSPE, San Raffaele Scientific Institute, Milan, Italy, ¹⁸Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA, ¹⁹Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA, ²⁰Institute of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg/Munich, Germany, ²¹Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany, ²²Department of Neuroscience, Neurology Section, University of Siena, Siena, Italy, ²³Department of Neuroscience, Ophthalmology and Genetics, University of Genoa, Genoa, Italy, ²⁴Department of Clinical Neurosciences, University of Palermo, Sicily, Italy, ²⁵Department of Neurological Sciences, Second University of

*To whom correspondence should be addressed at: National Institutes of Health, Building 35, Room 1A/1000, 35 Convent Drive, Bethesda, MD 20892-3720, USA. Tel: +1 3014517606; Fax: +1 3014517295; Email: traynorb@mail.nih.gov

†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.

Naples, Caserta, Italy, ²⁶Neurological Institute, Catholic University and I.CO.M.M. Association for ALS Research, Rome, Italy, ²⁷Department of Neuroscience, S. Agostino- Estense Hospital and University of Modena, Modena, Italy, ²⁸Department of Neurology, Center for Diagnosis and Cure of Rare Diseases, Bellaria Hospital, Bologna, Italy, ²⁹Department of Neuroscience, University of Pisa, Pisa, Italy, ³⁰University Department of Clinical Neurosciences, Institute of Neurology, University College London, London, ³¹Department of Neurology, Penn State College of Medicine, Hershey, PA, USA, ³²Department of Neurosurgery, Penn State College of Medicine, Hershey, PA, USA, ³³Department of Neurology, Cleveland Clinic, Cleveland, OH, USA, ³⁴Neurogenomics Division, Translational Genomics Institute (TGEN), Phoenix, AZ, USA, ³⁵Department of Neurodegenerative Disease, UCL Institute of Neurology, Queen Square, London, UK, ³⁶Institute of Clinical Neurobiology, University of Wuerzburg, Wuerzburg, Germany, ³⁷The ALS Association, Palm Harbor, FL, USA and ³⁸Department of Neurology, Johns Hopkins University, Baltimore, MD, USA

Received December 2, 2008; Revised and Accepted February 2, 2009

The cause of sporadic amyotrophic lateral sclerosis (ALS) is largely unknown, but genetic factors are thought to play a significant role in determining susceptibility to motor neuron degeneration. To identify genetic variants altering risk of ALS, we undertook a two-stage genome-wide association study (GWAS): we followed our initial GWAS of 545 066 SNPs in 553 individuals with ALS and 2338 controls by testing the 7600 most associated SNPs from the first stage in three independent cohorts consisting of 2160 cases and 3008 controls. None of the SNPs selected for replication exceeded the Bonferroni threshold for significance. The two most significantly associated SNPs, rs2708909 and rs2708851 [odds ratio (OR) = 1.17 and 1.18, and *P*-values = 6.98×10^{-7} and 1.16×10^{-6}], were located on chromosome 7p13.3 within a 175 kb linkage disequilibrium block containing the *SUNC1*, *HUS1* and *C7orf57* genes. These associations did not achieve genome-wide significance in the original cohort and failed to replicate in an additional independent cohort of 989 US cases and 327 controls (OR = 1.18 and 1.19, *P*-values = 0.08 and 0.06, respectively). Thus, we chose to cautiously interpret our data as hypothesis-generating requiring additional confirmation, especially as all previously reported loci for ALS have failed to replicate successfully. Indeed, the three loci (*FGGY*, *ITPR2* and *DPP6*) identified in previous GWAS of sporadic ALS were not significantly associated with disease in our study. Our findings suggest that ALS is more genetically and clinically heterogeneous than previously recognized. Genotype data from our study have been made available online to facilitate such future endeavors.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a rare and devastating neurodegenerative disease that predominantly affects motor neurons leading to progressive paralysis, and ultimately death from respiratory failure within 3–5 years of symptom onset. Approximately 5% of ALS is familial in nature, whereas the remaining 95% occurs sporadically throughout the community (1). Although the genetic causes of many monogenic familial forms of ALS have been described (2), the etiology of sporadic ALS is largely unknown. Familial aggregation studies, twin studies and epidemiological observations have suggested a substantial genetic contribution to disease risk (3,4). Recently, genome-wide association studies (GWAS) have putatively identified variants with moderate effects on the risk of developing ALS in the 1p32.1 region (*FGGY*) (5), in the 12p11 region (*ITPR2*) (6) and in the 7q36.2 region (*DPP6*) (7,8). However, these loci require replication in independent cohorts to confirm disease association, and, at most, account for only a fraction of the elevated risk of developing ALS, suggesting that additional genetic factors exist.

We conducted a two-stage GWAS to search for common variants with moderate risk (9,10). For the first stage, we

used 555 352 SNPs that extract information on 91% of common autosomal SNPs identified in the European populations based on the HapMap data [CEU, $r^2 > 0.8$, minor allele frequency (MAF) $> 5\%$] (10,11). These SNPs were genotyped in two independent cohorts of European origin consisting of 553 ALS cases and 2338 controls. For the second stage, we analyzed the 7600 SNPs that were most associated with the altered risk of disease in the initial genome-wide scan in an additional 2160 cases and 3008 controls. The large number of SNPs and samples genotyped in the second stage provided sufficient power to follow up on regions with moderate association in the initial genome-wide scan (threshold *P*-value for follow-up study < 0.005).

RESULTS

We conducted the initial genome-wide scan in a case-control cohort of 553 ALS cases and 2338 neurologically normal control of European ancestry. In the second stage, we genotyped 7600 of the most associated SNPs from the first stage in three additional replication cohorts totaling 2160 ALS cases, and compared this with data for the same 7600 SNPs in three control cohorts totaling 3008 samples. After quality

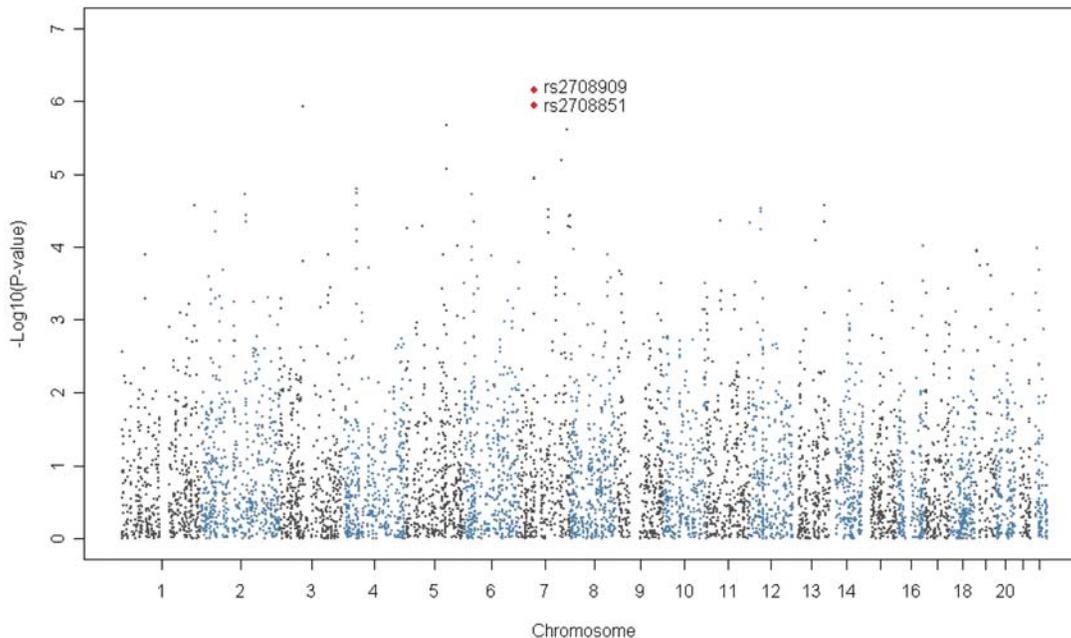


Figure 1. Association analysis of combined joints analysis in two-stage GWAS of ALS ($n = 2289$ ALS cases and 4532 controls) based on the logistic regression model correcting for age, gender and population. SNPs listed in Table 1 are represented by red dots.

control procedures, 6758 SNPs were available for analysis in a final combined stage 1 and stage 2 cohort of 2289 cases and 4532 controls. These SNPs covered 3152 distinct chromosomal regions defined by a maximal distance between two SNPs of <100 kb; 1745 regions contained only one SNP and 40 regions contained 10 or more SNPs. Of these regions, 94 had at least one SNP with an observed P -value $<10^{-3}$ (Fig. 1).

None of the SNPs tested in this study clearly achieved genome-wide significance after correction for multiple testing (see Supplementary Material, Table S3 for association results of all 6758 tested SNPs). The SNPs with the lowest P -values identified by our two-stage GWAS were located on chromosome 7p12.3 (Table 1), a region which has not been previously linked to the pathogenesis of ALS. The SNPs were located within a 175 kb linkage disequilibrium (LD) block containing three genes, *SUNCI*, *HUS1* and *C7orf57*. The strongest signal was observed for rs2708909 located in the third intron of the gene *SUNCI* (P -value = 6.98×10^{-7} in combined analysis, NM_152782.3), which encodes the 'SAD1 and UNC84 domain containing 1' protein. The second SNP, rs2708851, was in complete LD with rs2708909 ($r^2 = 0.97$; Fig. 2), and was located 22 kb upstream of *SUNCI* within intron 4 of *C7orf57* (NM_001100159.1). These SNPs did not exceed the threshold for genome-wide significance in the overall cohort, were only marginally associated with ALS risk when analyzed in the individual North American and Italian cohorts (P -values for rs2708909 = 5.40×10^{-5} and 0.0006, respectively), and were not associated in the German dataset (P -value for rs2708909 = 0.503), probably reflecting the smaller size of this cohort.

To further test the association with increased risk of disease, we genotyped rs2708851 and rs2708909 in a dataset of 989 US cases with ALS or other motor neuron diseases (MNDs)

and 327 neurologically normal US controls, all of whom were of non-Hispanic Caucasian ethnicity and had previously served in the US military (12,13). This sample set represented an independent sample, as none of these samples were included in the initial genome-wide stage or in the replication stage. Rs2708909 and rs2708851 failed to reach significance [P -value = 0.08 for rs2708909 and 0.06 for rs2708851 based on a logistic regression model correcting for age at onset and gender, odds ratio (OR) = 1.176 and 1.189, respectively], though the sample size was underpowered to detect moderate effect alleles (power to detect OR of 1.17 for a MAF of 0.45 = 41.1% at P -value of 0.05). The results are very similar when only patients diagnosed with definite or probable ALS were analyzed as cases (P -value for rs2708909 = 0.09; P -value for rs2708851 = 0.06). Furthermore, no evidence for association with the previously implicated SNPs in *ITPR2* and *DPP6* was found in this dataset (rs109260404: P -value = 0.62; rs2306677: P -value = 0.77) (14).

Rs2708909 and rs2708851 lie within a 175 kb region of LD (multiallelic $D' > 0.8$) on chromosome 7p12.3. Using our stage 1 datasets, we found that the HapMap CEU, the US and the Italian populations share an almost identical haplotype structure across this region (Supplementary Material, Fig. S2), and determined that seven SNPs (rs6955251, rs2686821, rs2686831, rs2708909, rs2708851, rs2307252 and rs2708912) account for 85% of the variation across the 175 kb region at an $r^2 > 0.5$. The first five of these markers had been genotyped as part of our stage 1 and stage 2 datasets. To investigate whether other SNPs in the same region were more significantly associated with altered risk of developing ALS, we analyzed genotype data for the two additional SNPs rs2307252 and rs2708912 for all stage 1 cases and controls (based on previous whole genome data, $n = 2521$), stage 2 controls (based on previous whole genome data, $n = 2548$) and 216 stage 2

Table 1. SNPs with a P -value $< 10^{-6}$ based on the logistic regression model of 2289 ALS cases and 4532 controls

	rs2708909	rs2708851
Chromosomal location	7p12.3	7p12.3
Position	48,018,204	48,052,327
Nearest gene (SNP type)	<i>SUNCI</i> (intronic)	<i>SUNCI</i> (intergenic)
Risk allele	G	C
Minor allele	T	T
Major allele	G	C
Minor allele frequency (cases)	0.449	0.451
Minor allele frequency (controls)	0.496	0.497
Allelic odds ratio (95% CI)	1.17 (1.11–1.23)	1.17 (1.11–1.23)
Population attributable risk (95% CI)	0.08 (0.05–0.12)	0.08 (0.05–0.12)
<i>All samples</i>		
P -value	6.98×10^{-7}	1.16×10^{-6}
Risk allele frequency (cases)	0.551	0.549
Risk allele frequency (controls)	0.504	0.503
Inter-SNP LD (r^2)	0.949	
<i>North American samples</i>		
P -value	5.40×10^{-5}	2.61×10^{-5}
Risk allele frequency (cases)	0.54	0.54
Risk allele frequency (controls)	0.488	0.512
Inter-SNP LD (r^2)	0.956	
<i>Italian samples</i>		
P -value	0.0006	0.0011
Risk allele frequency (cases)	0.590	0.589
Risk allele frequency (controls)	0.548	0.548
Inter-SNP LD (r^2)	0.927	
<i>German samples</i>		
P -value	0.503	0.611
Risk allele frequency (cases)	0.505	0.50
Risk allele frequency (controls)	0.489	0.488
<i>National Registry of Veterans with ALS</i>		
P -value	0.08	0.06
Risk allele frequency (cases)	0.499	0.499
Risk allele frequency (controls)	0.457	0.456
Inter-SNP LD (r^2)	0.961	

Position was based on NCBI Build 36.3; allele frequencies, odds ratios [with 95% confidence intervals (CI)] and population attributable risk were calculated using the combined stage 1 and stage 2 data; risk allele, the allele with higher frequency in cases compared with controls; minor allele, the allele with the higher frequency in controls; major allele, the allele with higher frequency in controls; P -values for *all samples* were based on the logistic regression model corrected for age, gender and population, and were based on the logistic regression model corrected for age and gender for the individual populations; all samples, 2289 ALS cases and 4532 controls included in the combined joint analysis; North American samples, 869 ALS cases and 2858 controls; Italian samples, 871 Italian ALS cases and 1190 Italian controls; German samples consisted of 549 ALS cases and 484 controls; National Registry of Veterans with ALS samples consisted of 989 ALS cases and 327 controls; inter-SNP LD (r^2), linkage disequilibrium between rs2708909 and rs2708851 measured as r^2 (Haploview 4.1) (36).

US cases (based on additional sequencing data). Neither rs2307252 nor rs2708912 achieved genome-wide significance (P -values = 0.47 and 0.16) based on this cohort of 753 cases and 4532 controls. Next, we applied imputation to our stage 1 data using MACH version 1.0, but none of the untyped SNPs in the region of 7p12.3 provided stronger evidence of association compared with rs2708909 and rs2708851 (Fig. 2).

Our two-stage GWAS identified several additional loci with P -values less than 10^{-3} representing hypotheses that may merit additional follow-up studies (Supplementary Material,

Table S4) (9,10). The three loci (*FGGY*, *ITPR2* and *DPP6*) identified in previous GWAS of sporadic ALS (5–8) did not alter risk of developing disease in either the combined case-control cohort or in the three individual populations examined in our study (Table 2).

Raw sample-level genotype data from the initial GWAS study (North American ALS cases, North American controls, Italian ALS cases and Italian controls from the Piemonte/Turin region) are available for download through the dbGAP portal (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000006.v1.p1). Individual genotype data for the CGEMS dataset are available for registered users through the CGEMS portal (<http://cgems.cancer.gov/data/>), whereas individual genotype data for the KORA cohort may be requested (http://epi.gsf.de/kora-gen/index_e.php).

DISCUSSION

Here we present the results of our two-stage GWAS involving 2713 cases and 5346 controls. This analysis was corrected based on the 531 661 autosomal SNPs genotyped in the initial whole genome scan, rather than the smaller number of SNPs followed up in the replication stage, as we wanted to be as conservative as possible. Our study represents the largest GWAS published to date in ALS, and the first to be sufficiently powered to reliably detect moderately associated SNPs in this relatively rare, fatal neurodegenerative disease.

Our study did not identify any SNPs that clearly exceeded the standard threshold for genome-wide significance (i.e. $< 10^{-7}$), and there is little or no overlap with the results of previously published studies (5–8). This contrasts with GWAS in other neurological diseases such as multiple sclerosis, where the most associated SNP in the HLA-DRA locus had a P -value $< 10^{-80}$ (15). One possible explanation for this lack of success may be that ALS is a more genetically and clinically heterogeneous disease than previously appreciated, which would significantly limit the power of GWAS. The identification of multiple different familial ALS genes, each involving disparate biological pathways, supports this notion (16). Alternatively, causative genetic factors may increase the risk of motor neuron degeneration by only a small amount (i.e. OR < 1.2) meaning that even larger GWAS involving 5000–10 000 will be required to reliably identify loci (17). Finally, we compensated for the relatively small number of cases in the first stage of our study by selecting several thousand SNPs for detailed follow-up. Although this approach is likely to be adequate for identifying alleles of moderate effects (i.e. OR of > 1.4), mild effect alleles could easily have failed to reach the threshold for inclusion in the replication stage.

Our study also failed to replicate the three loci (*FGGY*, *ITPR2* and *DPP6*) that had been previously published as been associated with increased risk of sporadic ALS (5–8). This finding agrees with data from the National Registry of Veterans with ALS, which also failed to replicate these loci in a cohort of 989 cases and 327 controls (14). There are several possible explanations for this finding. First, the lack of replication of these loci in the current study may be explained by the small number of SNPs selected from the

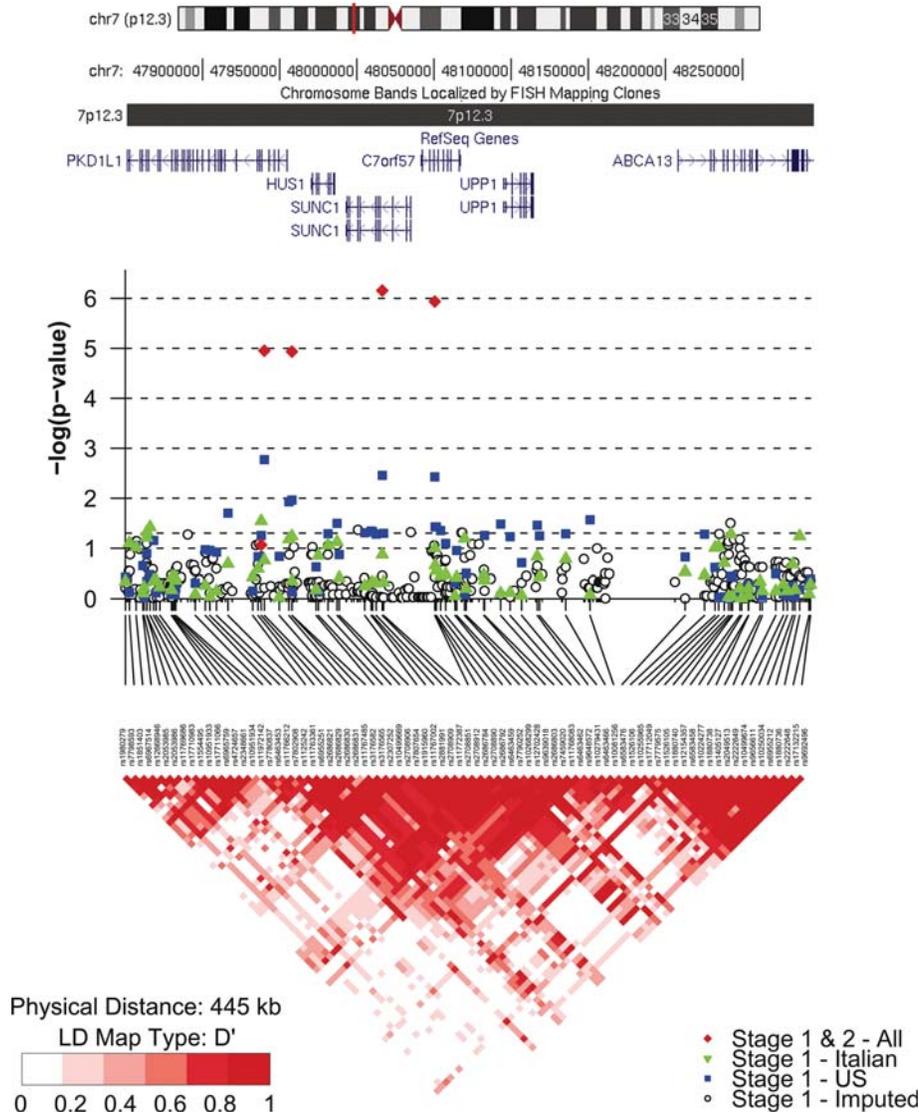


Figure 2. Location of the association signal and pairwise linkage disequilibrium (LD) surrounding the most associated SNPs on chromosome 7p12.3. LD pattern is depicted using stage 1 US data. Association signals are shown for all SNPs genotyped in (A) stage 1 US samples (blue squares, $n = 1065$); (B) stage 1 Italian samples (green triangles, $n = 1456$); (C) stage 2 all populations (orange circles, $n = 4300$), and the combined dataset (red diamonds, $n = 6821$). The most associated SNPs, rs2708909 and rs2708851, lie in or near gene *SUNCI*, and are in almost complete LD ($D' = 0.981$, $r^2 = 0.959$ based on stage 1 US data). Plots were produced using the *snp.plotter* package within R version 2.6.1.

initial genome-wide scans of the Dutch and TGEN studies for follow-up to confirm disease association. In the Dutch study of 461 cases and 450 controls, the 200 most associated SNPs were brought forward to the replication stages (6), whereas the TGEN study used a DNA pooling methodology involving 386 North American sporadic ALS cases and 542 controls to select 192 SNPs for individual-level genotyping (5). The several hundred thousand tests performed as part of any GWAS make it more likely that the most associated SNPs in the initial genome-wide scan represent false positive associations arising by chance ('winner's curse'). Indeed, previous two-stage GWAS studies have repeatedly shown that truly causative SNPs are often not ranked in the top 1000 SNPs in the initial genome-wide scan (10), which led us to select a large number of SNPs for replication in our stage 2 analysis.

Another possible explanation for the lack of replication of the *FGGY*, *ITPR2* and *DPP6* is that the initial Dutch and TGEN studies identified markers that are not in strong LD with the causal variant, leading to a false refutation in our study that was based on different populations (9).

The chromosome 7 risk variants putatively identified as hypotheses by our study were not associated with disease when analyzed in the individual German population included in the study. Although population-to-population variation in causative genes has been postulated for ALS (18,19), our findings are more likely to reflect the smaller number of the samples from the individual populations included in the study, and the consequent loss in power to detect moderate effect loci: the smallest German cohort of 549 cases and 484 controls had only 16.4% power to detect the *SUNCI* locus

Table 2. *P*-values from the two-stage GWAS for SNPs that have been previously associated with increased risk of developing ALS (5–7)

SNP	Chr	Gene	Risk allele (frequency affected)	<i>P</i> -value			
				All samples	US samples	Italian samples	German samples
rs6700125	1	<i>FGGY</i>	T (0.36)	0.08	0.23	0.55	0.35
rs2306677	12	<i>ITPR2</i>	T (0.11)	0.74	0.30	0.27	0.97
rs10260404	7	<i>DPP6</i>	C (0.39)	0.40	0.36	0.96	0.28

Chr, chromosome; risk allele, the allele with higher frequency in cases compared with controls (5–7); frequency affected, frequency of risk allele in 2289 ALS cases included in the current study; *P*-values for *all samples* were based on the logistic regression model corrected for age, gender and population, and were based on the logistic regression model corrected for age and gender for the individual populations; all samples, 2289 ALS cases and 4532 controls included in the combined joint analysis; North American samples, 869 ALS cases and 2858 controls; Italian samples, 871 Italian ALS cases and 1190 Italian controls; German samples consisted of 549 ALS cases and 484 controls.

(assuming an OR of 1.18 and a MAF of 0.45), whereas the larger North American dataset of 3727 samples had 55.1% power under the same parameters. Indeed, the putative association of rs2708909 and rs2708851 with ALS is only apparent in the combined analysis of 2289 cases and 4532 controls (power to detect *SUNCI* locus = 94.8%), emphasizing the necessity of using several thousand samples to detect variants that only moderately increase risk of developing sporadic ALS (20).

Even if we assume that the chromosome 7 variants are truly associated with ALS, we are left with the problem of determining which gene within this LD block is responsible for increased risk of disease. The location of the variant with the most significant *P*-value within the intron of *SUNCI* would suggest that this gene is the most likely candidate. Indeed, *SUNCI* encodes a 40.5 kDa nuclear envelope protein ‘Sad1 and UNC84 domain containing 1’ (21), and mutations in nuclear envelope proteins underlie a variety of neuromuscular diseases including Charcot-Marie-Tooth disease, type 2B1 (22) and spastin-associated hereditary spastic paraplegia (23). However, these biological hypotheses should be interpreted cautiously: although the gene lying closest to an associated SNP is generally considered to be the prime suspect in disease pathogenesis, a number of alternative pathogenic mechanisms must be considered: our own studies have shown that the associated SNP may ‘tag’ the true causative variant residing many kilobases distant in another gene; the associated SNP could affect expression of *cis* genes up to 100 Kb distant, or could act in *trans* to alter gene expression on other chromosomes (24); alternatively, the SNP could alter the function or tissue-specific expression of a previously unidentified microRNA or genetic element. Furthermore, despite the large number of samples analyzed in our study, replication of the locus in independent cohorts remains a necessity (9). The two SNPs reported in the current study did not achieve significant association with disease in a separate cohort of 221 Irish ALS cases and 211 neurologically normal controls, though the small size of this cohort precludes firm conclusions being drawn from this data (Irish data was not included in the current study as necessary covariates were not available from the investigators associated with the study) (7). Public release of raw genotype data is helpful in this regard, as it reduces the expense of future whole genome association studies and allows researchers to have greater confidence in the results of their association studies

by increasing sample size and power to accurately detect causative loci (25). Our initial public release of data established a powerful, unique resource for the ALS research community (25), and this data has been incorporated into all other ALS GWAS published to date (5–8). Coincident with publication, we have augmented this initiative with data from all 2713 ALS cases genotyped in the current study.

In summary, we present the results of our two-stage genome-wide association in a large cohort of sporadic ALS patients. None of the studied loci clearly achieved genome-wide level of significance, and none of the previously published loci were significantly associated with disease in our study. Though the data supporting an association of the chromosome 7p12.3 variants are suggestive, we chose to interpret these results cautiously as loci previously reported to be associated with increased risk of developing ALS have uniformly failed to replicate (26). Thus, these variants should be considered as hypothesis-generating that require additional replication to confirm or refute their veracity. The current lack of success of GWAS in sporadic ALS may indicate that the disease is more heterogeneous than previously recognized, and highlights the fact that even larger sample numbers will be required to definitively dissect the genetics underlying this fatal neurodegenerative disease.

MATERIALS AND METHODS

Initial genome-wide scans

We used HumanHap550 version one BeadChips (Illumina Inc., San Diego, CA, USA) to genotype 555 352 SNPs in (i) 276 North American ALS cases and 828 North American neurologically normal control samples obtained from the NINDS Neurogenetics repository at the Coriell Institute for Medical Research (NJ, USA). The initial part of this scan has been previously published, and the raw genotype data made publicly available (25); and (ii) 277 Italian ALS cases and 263 Italian control samples collected within the Turin/Piemonte area of Northern Italy. We also genotyped 561 466 SNPs in an additional cohort of 1247 control samples obtained from the InChianti study, a representative population-based cohort of older persons living in the Chianti geographic area (Tuscany, Central Italy) (27), using the HumanHap550 version three BeadChip (28). Analysis was confined to the 545 066 SNPs that are common to versions one and three

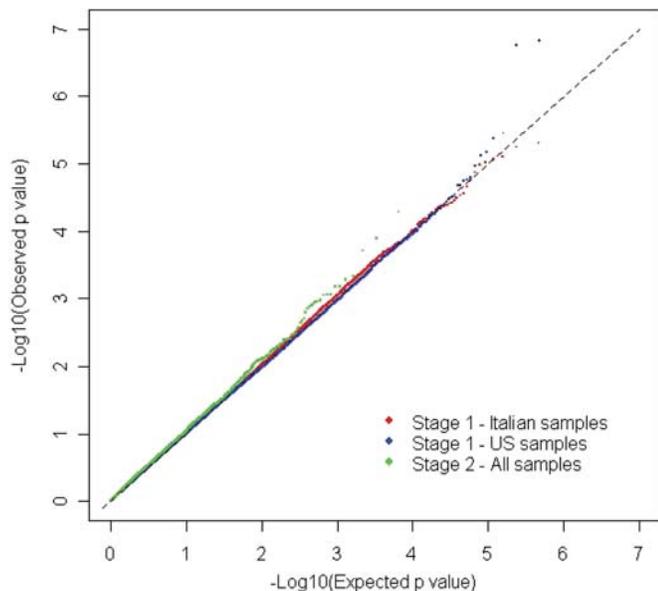


Figure 3. Q-Q plot based on (A) genomic-controlled Cochran-Armitage Trend test P -values for 474 554 SNPs analyzed in the stage 1 271 North American ALS cases and 794 North American controls (genomic inflation factor $\lambda = 1.002$); (B) genomic-controlled Cochran-Armitage Trend test P -values for 466 131 SNPs analyzed in the stage 1 266 Italian ALS cases and 1190 Italian controls (genomic inflation factor $\lambda = 1.147$) and (C) logistic regression P -values for 6758 SNPs analyzed in the stage 2 1752 ALS cases and 2548 controls.

HumanHap550 BeadChips. Demographics and clinical features of the case and control cohorts are shown in Supplementary Material, Table S1. All patients included in the initial and follow-up stages of the study had been diagnosed with probable, clinically probable, laboratory supported or definite sporadic ALS accorded to the El Escorial criteria (29). Only unrelated, non-Hispanic, white Caucasian individuals of European descent were included in the study.

Standard quality control procedures (e.g. exclusion of samples with low call rates, mismatch between gender according to phenotype data and gender defined by genotype, non-European ancestry, cryptic relatedness and incomplete phenotype data; and exclusion of SNPs with low call rates, Hardy-Weinberg equilibrium P -value < 0.001 and non-random missingness) were applied to the data (Supplementary Material, Fig. S1). After applying these filters, the cohorts included in the initial genome-wide scans consisted of (i) 271 North American sporadic ALS cases and 794 North American control individuals and (ii) 266 Italian sporadic ALS cases and 1190 Italian control samples. A total of 474,554 SNPs were available for association testing in the North American cohort, and 466 131 SNPs in the Italian cohort. Q-Q plots were prepared using R version 2.6.1 (2007, The R Foundation for Statistical Computing) based on genomically controlled association data (Fig. 3). Genomic inflation factor λ for the US cohort was 1.002 and for the Italian cohort was 1.147. The most significantly associated SNPs identified in the genome-wide scans of the North American and Italian cohorts are listed in Supplementary Material, Table S2.

Replication stage

A total of 7600 SNPs were selected on the basis of single-SNP association tests in the initial genome-wide scans. Of these, 2533 were selected as they were the most significantly associated SNPs in the North American cohort based on the Cochran-Armitage test; 2533 were the most significant SNPs in the Italian cohort and 2534 SNPs were chosen based on their disease association in the combined Italian and North American GWAS cohorts. These SNPs were genotyped using a custom-designed iSelect Infinium assay (Illumina) in three cohorts of sporadic ALS patients of European background totaling 2160 cases. The case dataset was compared with data for the same SNPs obtained from 3008 neurologically normal control individuals. The cases consisted of 963 North American sporadic ALS patients, 631 Italian sporadic ALS patients and 566 German individuals diagnosed with sporadic ALS. Population control data was obtained from: the Cancer Genetic Markers of Susceptibility study (CGEMS; $n = 2243$ North American control samples) (30); European Network for Genetic-Epidemiological Studies (HYPERGENES, $n = 275$ Italian control samples) and the Cooperative Health Research in the Region of Augsburg study (KORA; $n = 490$ German control samples) (31). These studies were approved by the appropriate institutional review boards.

After applying quality control filters to the replication data, 6758 SNPs were suitable for analysis in the replication cohort consisting of 1752 cases and 2548 controls (Supplementary Material, Fig. S1). Association analysis was carried out in two ways: for each population separately, and for the stage 1 and stage 2 replication data combined (32). The combined stage 1 and stage 2 samples consisted of 2289 ALS cases and 4532 controls, which yielded 94.8% power to detect loci with an OR of 1.17 and a MAF of 0.45 under the log-additive model assuming a two-sided α of 0.005 (threshold P -value for selection of SNPs for follow-up). The PLINK toolset (33) was used to test for association using logistic regression, adjusted for age, gender and population. This approach retains power to detect recessive or overdominant alleles at the cost of a small decrease in power relative to the Cochran-Armitage trend test (34) for the detection of alleles with multiplicative effect (10). Bonferroni correction for multiple testing yielded a threshold P -value of 10^{-7} ($\alpha = 0.05/531\,661$ autosomal SNPs genotyped in stage 1) (35).

Detailed descriptions of sample collection methodology and the quality control filters used in this study are available in Supplementary Material, Methods.

Additional replication dataset

Rs2708851 and rs2708909 were genotyped by TaqMan assays (ABI) in an additional dataset of 989 non-Hispanic Caucasian individuals diagnosed with ALS or other MNDs and 327 neurologically normal control samples. The case samples have been collected by the National Registry of Veterans with ALS (12), and the control samples by the GENEVA study (13). Of these 989 patients, 663 (67%) were diagnosed with definite or probable ALS by El Escorial criteria, 79 (8%) had possible ALS, 158 (16%) had progressive muscular

atrophy and the remaining 89 (9%) had primary lateral sclerosis or progressive bulbar palsy. These samples had not been genotyped in either the initial genome-wide stage or in the replication stage.

Haplotype block determination and imputation

Haploview v4.1 was used for assessment of LD (36). Haplotype blocks were defined using the Haploview v4.1 solid spine of LD method ($D' > 0.8$). SNPs selected to fine-map the haplotype block on chromosome 7p12.3 were genotyped by sequence analysis. PCRs were amplified using primers designed using Primer3-web v0.3.0 (<http://frodo.wi.mit.edu>) and FastStart PCR MasterMix polymerase (Roche Diagnostics Corp., IN, USA) were sequenced using BigDye terminator v3.1 sequencing chemistry, and run on an ABI3730xl (Applied Biosystems, CA, USA) genetic analyzer as per manufacturer's instructions. The sequences were analyzed with Sequencher software, version 4.2 (Genecodes, VA, USA). MACH version 1.0 software (<http://www.sph.umich.edu/csg/abecasis/MACH/download/>) (37) was used to estimate haplotypes, and map crossover and error rates using 100 iterations of the Markov chain Monte Carlo algorithm in subsets of 250 random samples from the stage 1 US and Italian cohorts. These estimates were combined with haplotypes from the HapMap CEU samples to infer genotypes in the region of interest for both cohorts using maximum likelihood estimates of genotypes present in HapMap samples (www.hapmap.org, release 21), but not in the Illumina data. After genotype imputation, the maximum likelihood genotypes for the stage 1 US and Italian cohorts were merged. Analyses were rerun, excluding imputed SNPs with r^2 values < 0.30 between imputed and known genotypes, and posterior probability averages < 0.80 for the most likely genotype imputed.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We would like to thank Katrina Gwinn, Larry Refolo and Roderick Corriveau of Coriell, The Italian Football League (FIGC), and Augustin Luna. DNA panels from the NINDS Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>) were used in this study, as well as clinical data. The submitters that contributed samples are acknowledged in detailed descriptions of each panel: NDPT002, NDPT006, NDPT008, NDPT011 to NDPT013, NDPT019 to NDPT030. We also thank the ALS Research Group and the patients who submitted their samples to the NINDS Repository and to other investigators.

Conflict of Interest statement. L.B. is Vice-President for Research, ALS Association and J.R. is Medical Director of the Packard Center for ALS Research. Other authors report no conflict of interest.

FUNDING

This work was supported by the Intramural Research Program of the National Institute on Aging (project Z01 AG000949-02), the National Institute of Neurological Disorders and Stroke, and the National Institute of Mental Health; the ALS Association; the Packard Center for ALS Research at Johns Hopkins; Istituto Superiore di Sanità (grant number 2005-10 to FL); PF A15 Regione Piemonte (to GR); and the Medical Research Council (to EF and DK). The MONICA/KORA Augsburg studies were financed by the Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany and supported by grants from the German Federal Ministry of Education and Research (BMBF). Part of this work was financed by the German National Genome Research Network (NGFN). The KORA research was supported within the Munich Center of Health Sciences (MC Health) as part of LMUinnovativ. We gratefully acknowledge support for the GENEVA study ('Genes and Environmental Exposures in Veterans with ALS') from the National Institutes of Health (grant number ES013244) and the ALS Association (grant number 1230). The National Registry of Veterans with ALS (grant number CSP #500A) and its DNA bank (grant number CSP #478) were supported by the Department of Veterans Affairs Cooperative Studies Program (CSP).

REFERENCES

- Chio', A., Traynor, B.J., Lombardo, F., Fimognari, M., Calvo, A., Ghiglione, P., Mutani, R. and Restagno, G. (2008) Prevalence of SOD1 mutations in the Italian ALS population. *Neurology*, **70**, 533–537.
- Valdmanis, P.N. and Rouleau, G.A. (2008) Genetics of familial amyotrophic lateral sclerosis. *Neurology*, **70**, 144–152.
- Graham, A.J., Macdonald, A.M. and Hawkes, C.H. (1997) British motor neuron disease twin study. *J. Neurol. Neurosurg. Psychiatry*, **62**, 562–569.
- Majoor-Krakauer, D., Ottman, R., Johnson, W.G. and Rowland, L.P. (1994) Familial aggregation of amyotrophic lateral sclerosis, dementia, and Parkinson's disease: evidence of shared genetic susceptibility. *Neurology*, **44**, 1872–1877.
- Dunckley, T., Huettelman, M.J., Craig, D.W., Pearson, J.V., Szelinger, S., Joshupura, K., Halperin, R.F., Stamper, C., Jensen, K.R., Letizia, D. *et al.* (2007) Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N. Engl. J. Med.*, **357**, 775–788.
- van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Andersen, P.M., Van Den, B.L., de Jong, S.W., van't, S.R., Birve, A. *et al.* (2007) ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet Neurol.*, **6**, 869–877.
- Cronin, S., Berger, S., Ding, J., Schymick, J.C., Washecka, N., Hernandez, D.G., Greenway, M.J., Bradley, D.G., Traynor, B.J. and Hardiman, O. (2008) A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum. Mol. Genet.*, **17**, 768–774.
- van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Van Den, B.L., de Jong, S.W., de, J.V., Baas, F., van't Slot, R. *et al.* (2008) Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.*, **40**, 29–31.
- Chanock, S.J., Manolio, T., Boehnke, M., Boerwinkle, E., Hunter, D.J., Thomas, G., Hirschhorn, J.N., Abecasis, G., Altshuler, D., Bailey-Wilson, J.E. *et al.* (2007) Replicating genotype-phenotype associations. *Nature*, **447**, 655–660.
- Thomas, G., Jacobs, K.B., Yeager, M., Kraft, P., Wacholder, S., Orr, N., Yu, K., Chatterjee, N., Welch, R., Hutchinson, A. *et al.* (2008) Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.*, **40**, 310–315.
- Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M. *et al.* (2007)

- A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851–861.
12. Allen, K.D., Kasarskis, E.J., Bedlack, R.S., Rozear, M.P., Morgenlander, J.C., Sabet, A., Sams, L., Lindquist, J.H., Harrelson, M.L., Coffman, C.J. *et al.* (2008) The National Registry of Veterans with amyotrophic lateral sclerosis. *Neuroepidemiology*, **30**, 180–190.
 13. Schmidt, S., Allen, K.D., Loiacono, V.T., Norman, B., Stanwyck, C.L., Nord, K.M., Williams, C.D., Kasarskis, E.J., Kamel, F., McGuire, V. *et al.* (2008) Genes and environmental exposures in veterans with amyotrophic lateral sclerosis: the GENEVA study. Rationale, study design and demographic characteristics. *Neuroepidemiology*, **30**, 191–204.
 14. Schmidt, S., Allen, K.D., Rimmler, J., Loiacono, V., Stanwyck, C., Williams, C., Munger, H., Hauser, M. and Oddone, E. (2008) Association of sporadic ALS with candidate genes and environmental risk factors: The Genes and Environmental Exposures in Veterans with ALS (GENEVA) Study. Abstract presented at 58th Annual Meeting, ASHG.
 15. Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I., Gabriel, S.B., Mirel, D.B., Iverson, A.J. *et al.* (2007) Risk alleles for multiple sclerosis identified by a genomewide study. *N. Engl. J. Med.*, **357**, 851–862.
 16. Pasinelli, P. and Brown, R.H. (2006) Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat. Rev. Neurosci.*, **7**, 710–723.
 17. Schymick, J.C., Talbot, K. and Traynor, B.J. (2007) Genetics of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.*, **16** (Spec No. 2), R233–R242.
 18. Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J. *et al.* (2006) ANG mutations segregate with familial and ‘sporadic’ amyotrophic lateral sclerosis. *Nat. Genet.*, **38**, 411–413.
 19. Lambrechts, D., Storkebaum, E., Morimoto, M., Del Favero, J., Desmet, F., Marklund, S.L., Wyns, S., Thijs, V., Andersson, J., van, M.I. *et al.* (2003) VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nat. Genet.*, **34**, 383–394.
 20. Wang, W.Y., Barratt, B.J., Clayton, D.G. and Todd, J.A. (2005) Genome-wide association studies: theoretical and practical concerns. *Nat. Rev. Genet.*, **6**, 109–118.
 21. Raff, J.W. (1999) The missing (L) UNC? *Curr. Biol.*, **9**, R708–R710.
 22. Sandre-Giovannoli, A., Chaouch, M., Kozlov, S., Vallat, J.M., Tazir, M., Kassouri, N., Szepietowski, P., Hammadouche, T., Vandenberghe, A., Stewart, C.L. *et al.* (2002) Homozygous defects in LMNA, encoding lamin A/C nuclear-envelope proteins, cause autosomal recessive axonal neuropathy in human (Charcot-Marie-Tooth disorder type 2) and mouse. *Am. J. Hum. Genet.*, **70**, 726–736.
 23. Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P. *et al.* (1999) Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nat. Genet.*, **23**, 296–303.
 24. Myers, A.J., Gibbs, J.R., Webster, J.A., Rohrer, K., Zhao, A., Marlowe, L., Kaleem, M., Leung, D., Bryden, L., Nath, P. *et al.* (2007) A survey of genetic human cortical gene expression. *Nat. Genet.*, **39**, 1494–1499.
 25. Schymick, J.C., Scholz, S.W., Fung, H.C., Britton, A., Arepalli, S., Gibbs, J.R., Lombardo, F., Matarin, M., Kasperaviciute, D., Crews, C. *et al.* (2007) Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls. *Lancet Neurol.*, **6**, 322–328.
 26. Garber, K. (2008) Genetics. The elusive ALS genes. *Science*, **319**, 20.
 27. Ferrucci, L., Bandinelli, S., Benvenuti, E., Di Iorio, A., Macchi, C., Harris, T.B. and Guralnik, J.M. (2000) Subsystems contributing to the decline in ability to walk: bridging the gap between epidemiology and geriatric practice in the InCHIANTI study. *J. Am. Geriatr. Soc.*, **48**, 1618–1625.
 28. Melzer, D., Perry, J.R., Hernandez, D., Corsi, A.M., Stevens, K., Rafferty, I., Lauretani, F., Murray, A., Gibbs, J.R., Paolisso, G. *et al.* (2008) A genome-wide association study identifies protein quantitative trait loci (pQTLs). *PLoS Genet.*, **4**, e1000072.
 29. Brooks, B.R., Miller, R.G., Swash, M. and Munsat, T.L. (2000) El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph. Lateral Scler. Other Motor Neuron Disord.*, **1**, 293–299.
 30. Yeager, M., Orr, N., Hayes, R.B., Jacobs, K.B., Kraft, P., Wacholder, S., Minichiello, M.J., Fearnhead, P., Yu, K., Chatterjee, N. *et al.* (2007) Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat. Genet.*, **39**, 645–649.
 31. Wichmann, H.E., Gieger, C. and Illig, T. (2005) KORA-gen – resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen*, **67** (suppl. 1), S26–S30.
 32. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209–213.
 33. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 34. Margolin, B.H. (1988) Test for trend in proportions. In Klotz, S. and Johnson, N.L. (eds), *Encyclopedia of Statistical Sciences*. John Wiley & Sons, Inc., New York, pp. 334–336.
 35. Rice, T.K., Schork, N.J. and Rao, D.C. (2008) Methods for handling multiple testing. *Adv. Genet.*, **60**, 293–308.
 36. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.
 37. Li, Y. and Abecasis, G.R. (2006) Rapid haplotype reconstruction and missing genotype inference. *Am. J. Hum. Genet.*, **S79**, 2290.