doi:10.1093/brain/awm055 *Brain* (2007), **130**, 2292–2301

Large-scale pathways-based association study in amyotrophic lateral sclerosis

Dalia Kasperavičiūtė,[1] Mike E. Weale,[2] Kevin V. Shianna,[3] Gareth T. Banks,[1] Claire L. Simpson,[4] Valerie K. Hansen,[4] Martin R. Turner,[4] Christopher E. Shaw,[4] Ammar Al-Chalabi,[4] Hardev S. Pall,[5,6] Emily F. Goodall,[5] Karen E. Morrison,[5,6] Richard W. Orrell,[7] Marcus Beck,[8] Sibylle Jablonka,[9] Michael Sendtner,[9] Alice Brockington,[10] Paul G. Ince,[10] Judith Hartley,[10] Hannah Nixon,[10] Pamela J. Shaw,[10] Giampietro Schiavo,[11] Nicholas W. Wood,[12] David B. Goldstein[2] and Elizabeth M. C. Fisher[1]

[1]Department of Neurodegenerative Disease, Institute of Neurology, University College London, London, UK, [2]IGSP Center for Population Genomics and Pharmacogenetics, Duke University, [3]IGSP Center for Applied Genomics and Technology, Duke University, Durham, NC, USA, [4]MRC Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry, Department of Neurology, London, [5]Department of Clinical Neurosciences, Division of Neuroscience, The Medical School, University of Birmingham, [6]Neuroscience Centre, Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, [7]Department of Clinical Neurosciences, Royal Free and University College Medical School, University College London, London, UK, [8]Department of Neurology, University of Wuerzburg, [9]Institute of Clinical Neurobiology, University of Wuerzburg, Wuerzburg, Germany, [10]Academic Neurology Unit, Section of Neuroscience, University of Sheffield Medical School, Sheffield, [11]Molecular NeuroPathobiology Laboratory, Cancer Research UK London Research Institute, and [12]Department of Molecular Neuroscience, Institute of Neurology, London, UK

Correspondence to: Dr D. Kasperavičiūtė, Department of Neurodegenerative Disease, Institute of Neurology, University College London, Queen Square, London WC1N 3BG, UK
E-mail: d.kasperaviciute@prion.ucl.ac.uk

Sporadic amyotrophic lateral sclerosis (ALS), a devastating neurodegenerative disease, most likely results from complex genetic and environmental interactions. Although a number of association studies have been performed in an effort to find genetic components of sporadic ALS, most of them resulted in inconsistent findings due to a small number of genes investigated in relatively small sample sizes, while the replication of results was rarely attempted. Defects in retrograde axonal transport, vesicle trafficking and xenobiotic metabolism have been implicated in neurodegeneration and motor neuron death both in human disease and animal models. To assess the role of common genetic variation in these pathways in susceptibility to sporadic ALS, we performed a pathway-based candidate gene case-control association study with replication. Furthermore, we determined reliability of whole genome amplified DNA in a large-scale association study. In the first stage of the study, 1277 putative functional and tagging SNPs in 134 genes spanning 8.7 Mb were genotyped in 822 British sporadic ALS patients and 872 controls using whole genome amplified DNA. To detect variants with modest effect size and discriminate among false positive findings 19 SNPs showing a trend of association in the initial screen were genotyped in a replication sample of 580 German sporadic ALS patients and 361 controls. We did not detect strong evidence of association with any of the genes investigated in the discovery sample (lowest uncorrected *P*-value 0.00037, lowest permutation corrected *P*-value 0.353). None of the suggestive associations was replicated in a second sample, further excluding variants with moderate effect size. We conclude that common variation in the investigated pathways is unlikely to have a major effect on susceptibility to sporadic ALS. The genotyping efficiency was only slightly decreased (~1%) and genotyping quality was not affected using whole genome amplified DNA. It is reliable for large scale genotyping studies of diseases such as ALS, where DNA sample collections are limited because of low disease prevalence and short survival time.

Keywords: amyotrophic lateral sclerosis; genetic association; axonal transport; whole genome amplification

Abbreviations: ALS = Amyotrophic lateral sclerosis; SMN = Survival motor neuron

Received December 11, 2006. Revised January 25, 2007. Accepted February 26, 2007. Advance Access publication April 17, 2007

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease characterized by progressive muscle weakness and wasting with combined upper and lower motor neuron loss. It is the most common motor neuron disease in the developed world with a lifetime risk of 1 : 600 to 1 : 1000 (Pasinelli and Brown, 2006). About 10% of ALS cases are familial mostly with autosomal dominant inheritance (Pasinelli and Brown, 2006); the remaining cases, often referred to as sporadic, most likely result from complex genetic and environmental interactions. Mutations in the superoxide dismutase 1 (*SOD1*) gene are identified in up to 20% of familial (Rosen, 1993) and in 3–7% of sporadic ALS patients (Jones *et al*., 1994, 1995; Jackson *et al*., 1997) and are the most frequent known cause of ALS. A few other genes, for example senataxin (*SETX*) (Chance *et al*., 1998; Chen *et al*., 2004), VAMP (vesicle-associated membrane protein)-associated protein B (*VAPB*) (Nishimura *et al*., 2004), and alsin (*ALS2*) (Yang *et al*., 2001; Hadano *et al*., 2001), have been shown to be mutated in rare forms of familial ALS and other forms of motor neuron degeneration (Pasinelli and Brown, 2006; James and Talbot, 2006); however, the aetiology of more than 98% of ALS cases remains unclear. Given that most of our understanding of disease pathogenesis at a molecular level comes from research on SOD1-related ALS, which comprises only ~2% of all cases, it is crucial to identify other disease risk factors, both genetic and environmental.

To date, no genetic risk factors have been unequivocally shown to be associated with sporadic ALS, the best replicated findings so far being regulatory polymorphisms in vascular endothelial growth factor (*VEGF*) (Lambrechts *et al*., 2003), copy number variation at the Survival Motor Neuron (*SMN*) locus (Corcia *et al*., 2002, 2006; Veldink *et al*., 2005) and differences in tail lengths in the heavy chain neurofilament gene (*NEFH*) (Al-Chalabi *et al*., 1999). One reason for this lack of associations is highlighted by Simpson and Al-Chalabi (2006) in that most of the previous genetic studies of sporadic ALS have been limited to assessing a small number of genes within patient and control groups of relatively small sample sizes—generally a few hundred patients and controls—resulting in inconsistent findings. For the same reason, replications of association are only very rarely reported or attempted. Until recently, the difficulties in carrying out such studies have been the lack of availability of human genome variation data and the available platforms for high throughput analysis of genetic variation. Within the last 2 years, new data on common human genetic variation have been published by the International HapMap project (Altshuler *et al*., 2005) and new and more reliable and robust genotyping platforms have become available, both of which greatly enhance our ability to carry out large-scale association studies. The HapMap project has facilitated high throughput analysis of human variation by providing the correlational structure of single nucleotide polymorphisms (SNPs) which enables us to select small set of tagging SNPs for genotyping to capture the most common variants in large portions of the human genome (de Bakker *et al*., 2005).

Given these new developments, we set out to undertake an integrated genetic association study in ALS, for the first time focusing on specific pathways/protein complexes of interest, as determined by published data implicating members of the pathway in ALS specifically or in other related forms of neurodegeneration. The three pathways/protein complexes we studied were those involved in (i) axonal transport, specifically retrograde axonal transport, (ii) vesicle trafficking and (iii) xenobiotic metabolism.

Axonal retrograde transport: This form of axonal transport is driven by cytoplasmic dynein, a multi-subunit motor complex moving towards the minus end of microtubules. Cytoplasmic dynein interacts with dynactin, which acts as a cargo adaptor and affects motor processivity (Pfister *et al*., 2006; Duncan and Goldstein, 2006). Defects in axonal transport have been implicated in motor neuron degeneration both in human disease and in mouse models. Transgenic mice overexpressing human mutant *SOD1* that model ALS have been shown to have slower retrograde axonal transport (Murakami *et al*., 2001; Kieran *et al*., 2005) and mutant SOD1 is known to disrupt dynein localization (Murakami *et al*., 2001; Ligon *et al*., 2005). Mutations in the cytoplasmic dynein 1 heavy chain 1 (*Dync1h1*) gene in mice result in slower retrograde axonal transport and death of motor neurons (Hafezparast *et al*., 2003), while interactions of the same mutant *Dync1h1* alleles with mutant SOD1 increase lifespan of ALS mice, and these double mutant mice have been shown to have increased rates of retrograde axonal transport (Kieran *et al*., 2005). A mutation in the dynactin p150 subunit was identified in a family with a slowly progressive lower motor neuron syndrome (Puls *et al*., 2003, 2005) and other possible mutations have been identified in ALS patients (Munch *et al*., 2004), while overexpression of dynamitin, another subunit of dynactin, in mice causes disruption of the dynein–dynactin complex and late onset motor neuron disease (LaMonte *et al*., 2002). Thus disruption of retrograde axonal transport is implicated in specific degeneration of motor neurons, and in some kind of interaction with mutant SOD1.

Vesicle trafficking: Defects in vesicular trafficking are known to lead to death of motor neurons. The *wobbler* mouse model for example has a mutation in the vesicular sorting protein *Vps54* (Schmitt-John *et al*., 2005), while a mutation in *CHMP2B*, a component of the endosomal secretory complex required for transport (ESCRTIII), may result in frontotemporal dementia in a human pedigree (Skibinski *et al*., 2005), a disease with considerable overlap with ALS; two possible mutations in *CHMP2B* have been identified in individuals with ALS (Parkinson *et al*., 2006).

Xenobiotic metabolism: Epidemiological studies have shown association of ALS with exposure to environmental toxins, pesticides and heavy-metals (reviewed in Nelson and

McGuire, 2006), although the results of many published studies are inconclusive. The exposure to such agents is modulated by the xenobiotic metabolizing enzymes of an individual, and enzyme activity is in part determined by genetic variation in the genes encoding these enzymes. Therefore we hypothesize that susceptibility to ALS could be influenced by genetic variation in genes in the xenobiotic metabolism pathways.

In summary, here we elucidate the role of common genetic variation in retrograde axonal transport, vesicle trafficking and major xenobiotic metabolism genes in susceptibility to sporadic ALS in the most powerful association study of ALS so far. We are specifically studying individual pathways in this investigation. We comprehensively screen variation in 134 genes in the largest collection of sporadic ALS patients genotyped in a single study to date and validate the results of association analysis by genotyping SNPs showing a trend of association in a replication sample of ALS patients from Germany.

In addition, by performing the initial screen entirely on whole genome amplified DNA we validate its use on a large scale. The reduction in the amount of DNA required for genotyping may significantly ameliorate the use of existing old DNA collections and enhance collaborations studying late onset neurodegenerative diseases where sample collections are limited because of low disease prevalence and short survival time. The use of whole genome amplified DNA until now has been limited to supplementing the studies performed on genomic DNA with small numbers of samples for which not enough DNA was available, and to the best of our knowledge this is the largest genome screen performed on whole genome amplified DNA.

Subjects and methods

Study plan and subjects

We used a two-stage study design with existing collections of sporadic ALS cases and controls from two populations. In the first, or discovery sample, we genotyped 822 British sporadic ALS patients and 872 control DNA samples collected at out-patient clinics at the Motor Neurone Disease Care and Research Centre, Queen Elizabeth Hospital, Birmingham, UK (167 definite or probable ALS patients according to El Escorial World Federation criteria with unknown *SOD1* mutation status and 145 controls), King's Motor Nerve Clinic, London, UK (258 definite or probable ALS patients according to El Escorial criteria with no *SOD1* mutations and 245 controls), the Newcastle and Sheffield MND Centres, UK (84 ALS patients neuropathologically confirmed at autopsy, 280 definite or probable ALS patients according to El Escorial criteria, 33 patients with clinical variants of ALS including primary lateral sclerosis and progressive muscular atrophy with unknown *SOD1* mutation status and 313 controls) and National Blood Transfusion Service, UK (169 controls). The second, or replication sample, consisted of 580 German sporadic definite or probable ALS patients according to El Escorial criteria with unknown *SOD1* mutation status, collected at the Motor Neuron Research Clinic, Wuerzburg, Germany and 361 controls, collected at the Department of Transfusion Medicine and Immunohematology, University of Wuerzburg, Germany. None of the patients had a known family history of ALS. The sample characteristics are given in Table 1. DNA samples were extracted from blood using standard methods. Participants of the study signed informed consent and the study was approved by the National Hospital for Neurology and Neurosurgery and Institute of Neurology Joint Research Ethics Committee.

A set of SNPs covering most of the common variants in candidate gene regions was genotyped in the discovery sample and statistical analysis performed to look for evidence for association with susceptibility to ALS. To detect modest effect size variants and to discriminate among false positive findings, a subset of SNPs that showed a trend for association in the discovery sample was genotyped in the replication sample.

Candidate gene and SNP selection

The candidate genes were selected following a detailed review of the literature on ALS-linked pathways and our experimental data (DK, GS and EMCF, unpublished data). We included genes encoding all known subunits of the dynein–dynactin complex, genes regulating its activity, binding to cargoes, microtubules and other interacting proteins; all known subunits of ESCRT complexes known to be involved in vesicle trafficking; key enzymes involved in pesticide metabolism and their targets. The complete gene list is shown in Supplementary Table 1. The total size of genome sequence investigated was 8.7 Mb.

SNPs were selected in candidate gene regions based on positions of RefSeq genes in the UCSC genome browser server (http://genome.ucsc.edu) hg16 assembly, adding 10 000 bp to the most 5′- and 3′-extent of the gene. As direct typing of causal variants is more powerful in association studies, we enriched the set of markers selected to genotype with the SNPs with predicted functionality. First, selected regions were screened for SNPs with predicted functionality using TAMAL software (Hemminger *et al.*, 2006). The SNPs were prioritized for genotyping if they had frequency data in any major databases (dbSNP, HapMap, Perlegen and Affymetrix) and met one of the following criteria: were in coding regions, altered intronic splice sites, were in predicted promoters, in regions with predicted regulatory potential, in predicted transcription factor binding sites, in regions with conservation scores $\geqslant$99th percentile genome-wide for human-chimp-rat-mouse-chicken alignment or in miRNAs and their 3′UTR targets. Functional SNPs for cytochrome P450 genes were selected from the CYP450 database (http://www.imm.ki.se/CYPalleles/). Secondly, tagging SNPs were selected based on HapMap PhaseII data (release 19) using Tagger software

Table 1 Characteristics of genotyped ALS patient and control samples

	British	German
ALS patients	822	579
Sex (males/females)	500/322	348/231
Type of onset (limb/bulbar/mixed or undetermined)	513/234/75	444/132/3
Mean age of onset (range)	59 (20–87)	58 (16–85)
Controls	872	361
Sex (males/females/unknown)	430/442/0	190/136/35
Mean age at sample collection (range)	52 (17–81)	32 (18–65)

(de Bakker *et al.*, 2005). Since the majority of ALS patients were of 'white' ethnicity, only data on the HapMap CEU population were used in selection of tagging SNPs. Where possible, putative functional SNPs were used as tagging SNPs, and SNPs with higher predicted genotyping scores from Illumina were prioritized in selection—only SNPs with Illumina genotype scores >0.6 were included in the assay. In total, 1279 tagging SNPs were selected to capture variation of polymorphisms with minor allele frequencies >5% in the HapMap CEU population with mean maximum pairwise r^2 between tagging SNP and ungenotyped SNP of 0.90. Eighty-three percent of all alleles with minor allele frequency of >0.05 were captured with $r^2 > 0.8$ and 95.3% with $r^2 > 0.5$. 157 SNPs not genotyped in the HapMap project were included because of predicted functionality. One hundred neutral SNPs were selected to control for population stratification. They were selected to have minor allele frequencies of >5% in the HapMap CEU population and to be in more than 50 kb distance from any RefSeq gene, known gene or RNA gene in UCSC genome hg17 assembly and not to be in the regions with predicted functionality (as earlier).

Genotyping and quality control in the discovery sample

British ALS patient and control samples were genotyped by 1536-plex GoldenGate assay on an Illumina BeadArray station using whole genome amplified DNA. Whole genome amplification was performed using a Qiagen Repli-g Midi kit according to manufacturer's instructions using 100 ng of input genomic DNA per reaction. The yields of amplified DNA were quantified using a Picogreen assay (Molecular Probes, Inc) and concentrations adjusted to 100 ng/μl. The sex of the samples was verified using the amelogenin locus as a marker. We found 2% sex mismatches and by typing additional sex-linked markers all of them were confirmed to be errors in labelling the original genomic DNA samples, therefore these samples were excluded from the study. Three percent of whole genome amplified samples failed in PCRs during sex testing and were removed from further study.

Genotyping quality control was ensured by (i) mixing case and control samples on the same plates and genotyping blind to the affection status; (ii) each plate contained five duplicate samples, including one genomic DNA duplicate; (iii) three whole genome amplified DNA samples from the Centre d'Etude du Polymorphisme Humaine (CEPH) were genotyped and results compared with HapMap data; (iv) SNPs that failed in more than 1% of the samples were removed to avoid non-random missing data; and (v) SNPs showing departures from Hardy–Weinberg equilibrium in controls were re-evaluated after analysis to check for possible genotyping errors and differences in genotyping performance between patient and control samples, such as separate clustering of patient and control samples and differences in intensity values.

Statistical analysis

The association was assessed by comparing genotype and allele frequencies between affected and unaffected individuals. For the genotypic test, a contingency table was made up for each SNP consisting of the three genotype categories on one axis and the phenotype categories on the other axis. The table was then analysed using an extension of Fisher's Exact Test to R*C tables, using a network algorithm developed by Mehta and Patel (1986) and implemented using R software. To assess family-wise significance of *P*-values, permutation analysis was performed using routines written in R. False discovery rate analyses were performed using the Benjamini and Hochberg linear step-up procedure (Benjamini and Hochberg, 1995). We assumed the proportion of true null hypotheses was close to 1, and that tests were positive regression dependent (Benjamini and Yekutieli, 2001). The effect of hidden population substructure on association statistics was assessed using the GCF Genomic Control method of Devlin *et al.* (2004). Tests of sets of *P*-values against their uniform expectation under the null, assuming independence of tests, were performed using Fisher's method for combining *P*-values.

Replication analysis

Permutations were performed to assess the significance of the association results in the discovery sample and prioritize SNPs for replication. In addition, all SNPs with uncorrected *P*-values below 0.01 in the genotypic test in the discovery sample were genotyped in the replication sample, excluding SNPs in high linkage disequilibrium with each other and SNPs from the genomic control set. We also genotyped SNPs with *P*-values between 0.01 and 0.05 if they had a strong prediction of functionality according to the following criteria: were in coding regions, within 6 bp from exon–intron boundaries in fastDB database (de la Grange *et al.*, 2005), in 3'UTRs or experimentally proven promoters.

Blinded genotyping was performed using Taqman on a 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA) using genomic DNA. As a quality control measure duplicates (100% agreement) and water blanks were used. Association statistics were calculated as earlier.

Results

Genotyping of whole genome amplified DNA

In total 1831 samples (95.4%) were genotyped successfully on the Illumina BeadArray system, including 75 whole genome amplified DNA blind duplicates and 17 genomic DNA—whole genome amplified DNA blind duplicates. 1372 SNPs (89%) passed quality control criteria for genotyping. We did not observe any differences between genomic DNA and whole genome amplified sample duplicates in genotyping performance.

In total 2 510 789 usable genotypes were produced including 126 141 blind duplicates, of which 23 311 were genomic DNA—whole genome amplified DNA duplicates. Two mismatched duplicated genotypes were detected, both of them in the same whole genome amplified duplicate sample, indicating overall error rate 1.59×10^{-5}. This also suggests that errors are sample specific, but not likely related to locus amplification bias during whole genome amplification, as these duplicate samples were the products of the same reaction. All genomic DNA—whole genome amplified DNA duplicate genotypes matched exactly. Genotypes of three whole genome amplified CEPH samples were compared to genotypes from the HapMap project data and one mismatch in 3706 compared genotypes was found, which indicates an error rate of 0.00027.

Marker coverage

1372 SNPs were genotyped successfully, including:

(1) 1156 (90% of attempted) tagging SNPs in 134 candidate gene regions, which captured the variation of 5283 SNPs with minor allele frequencies of greater than 5% in the HapMap CEU sample with a mean maximum r^2 0.865. 76.4% of alleles were captured with $r^2 > 0.8$ and 92.3% with $r^2 > 0.5$. 301 of the tagging SNPs had predicted functionality.
(2) 121 (77% of attempted) putative functional SNPs. These SNPs were genotyped because of predicted functionality, but either had minor allele frequency <5%, therefore could not be used as tagging SNPs, or had not been genotyped in the HapMap project therefore there was no prior information of their tagging properties. Twenty of these SNPs were monomorphic in our sample. We observed the largest dropout in this set, despite the fact that all these SNPs were validated by frequency in at least one database (HapMap, dbSNP, Perlegen or Affymetrix).
(3) Ninety-five (95% of attempted) SNPs from genomic control set. All these SNPs had experimentally validated Illumina assays (scores 1.1), therefore the genotyping performance, as expected, was the best.

Hardy–Weinberg equilibrium and population substructure

We examined the distribution of Hardy–Weinberg equilibrium *P*-values against the null expectation under both Hardy–Weinberg equilibrium and linkage equilibrium. The distribution of Hardy–Weinberg equilibrium *P*-values showed a significant difference from uniform distribution in controls ($P = 0.03$), but not in patients ($P = 0.13$). However, after careful investigation we did not detect any differences in genotyping quality between patient and control samples, and therefore we believe the excess of deviations from Hardy–Weinberg equilibrium may indicate population substructure in the control sample rather than errors in genotyping. We note that the two SNPs with the lowest *P*-values in deviation from Hardy–Weinberg equilibrium were in the same gene (*TRA1*) and in linkage disequilibrium (D' = 1), suggesting this is a real feature of this locus. If these two SNPs are removed, the *P*-value distribution is no longer significantly different from uniform ($P = 0.16$). Modest population substructure was also suggested by a marginal excess of low *P*-values in association statistics in the genomic control set, where the mean chi-square statistic over all loci differed significantly from null expectation ($P = 0.04$). Hidden population substructure may be the cause of false positive results if disease frequency is different between subpopulations as well as mask real associations. One method to deal with the former is Genomic Control (Devlin *et al.*, 2004), which provides a simple correction to the single-marker chi-square statistic for association that is applied uniformly across all SNPs. None of the SNPs tested achieved formal significance, even before genomic control correction (see later), and the ranked order of *P*-values, used for nominating SNPs for replication analysis, is unaffected by genomic control correction (Supplementary Table 2).

Association analysis in the discovery sample

We ranked SNPs according to their genotypic *P*-values for association. There were 17 SNPs with an uncorrected *P*-value of <0.01 for the genotypic test (Tables 2 and 3, Supplementary Table 2). We investigated the significance of these findings using both Family-Wise Error Rate (FWER) and False Discovery Rate (FDR) methods (Table 2 and Supplementary Table 2). To account for dependency among tests due to linkage disequilibrium, we used permutations to estimate multiplicity-adjusted (family-wise) *P*-values. None of the *P*-values reached significance after 1000 permutations, the lowest family-wise adjusted *P*-value being 0.353 (rs7961369 in *BICD1*). We also calculated multiplicity-adjusted *q*-values, defined as the minimum FDR required to admit the *k*th ranked SNP into the pool of declared hits. The minimum expected FDR required to admit even one SNP under this method was 0.284. Together, these results do not provide significant support for association between the analysed SNPs and susceptibility to ALS, over the alternative hypothesis of no associated SNPs.

We repeated the analysis excluding the 33 samples from patients with the primary muscular atrophy and primary lateral sclerosis clinical subtypes of ALS and who did not therefore strictly meet the El Escorial criteria for probable or definite ALS. These subtypes were included in the initial analysis because of the strong clinical suspicion that they are part of the same disease process as classical ALS. The results of both analyses highly correlated with r^2 of 0.97. Importantly, the 20 SNPs with lowest *P*-values, including all SNPs with $P < 0.01$ were the same in both analyses. Primary muscular atrophy and primary lateral sclerosis may be considered to be the same pathological process as ALS, therefore we chose to use the results of analysis of the full dataset for SNP selection for replication.

Association analysis in the replication sample

Even though we did not detect unequivocal evidence of a significant association in the initial screen, a Q–Q plot of log observed *P*-values against expected indicated some excess of SNPs with low *P*-values in our dataset. It could be indicative of the presence of alleles with weak and moderate effect size which our study was underpowered to detect, even though the difference did not reach formal significance ($P = 0.12$). To discriminate between false positives and alleles with moderate effect size, we genotyped 19 SNPs in a replication sample of 580 German sporadic ALS patients

Table 2 Characteristics of SNPs genotyped in discovery and replication samples and association statistics, ranked according to *P*-values in the discovery (British) sample.

SNP ID	Gene	Location in gene	Function	Genotype counts in British patients	Genotype counts in British controls	Unadjusted *P*-value in British	FDR *q*-value in British	HWE *P*-value in British controls	Genotype counts in German patients	Genotype counts in German controls	Unadjusted *P*-value in Germans	HWE *P*-value in German controls	Joint *P*-value
rs7961369	*BICD1*	Intron	None	175/387/260	196/474/202	0.00037	0.284	0.010	144/259/165	82/162/103	0.84	0.242	0.0031
rs17488186	*CAPZA3*	Upstream	Promoter (predicted)	0/11/811	0/35/837	0.00081	0.301	0.545	0/17/564	0/10/351	1	0.790	0.012
rs17191358	*ANXA2*	Upstream	None	13/141/668	17/212/643	0.00089	0.301	0.922	8/131/438	7/86/263	0.64	0.992	0.0060
rs37599111	*ANXA2*	Intron	None	58/283/480	84/352/436	0.0016	0.324	0.295	46/193/245	28/155/155	0.23	0.212	0.0034
rs9926649	*DYNC1LI2*	Downstream	None	7/69/746	0/99/773	0.0017	0.324	0.076	0/48/525	1/31/321	0.48	0.785	0.025
rs2356606	*DISC1*	Intron	None	60/262/500	32/311/529	0.0023	0.350	0.095	23/209/343	18/117/220	0.47	0.634	0.09
rs6727909	*VPS54*	Intron	None	7/68/747	0/94/778	0.0034	0.460	0.093	1/70/502	0/34/321	0.32	0.343	0.019
rs2306419	*ANXA5*	Intron	None	2/122/698	3/86/783	0.0043	0.524	0.698	4/72/496	4/64/288	0.053	0.834	0.38
rs2171209	*VIL2*	Intron	None	46/210/566	34/281/557	0.0048	0.537	0.846	34/185/351	20/130/205	0.43	0.918	0.0096
rs246129	*VPS4A*	Upstream	None	106/323/393	86/404/382	0.0074	0.692	0.163	49/220/212	30/174/144	0.44	0.025	0.0087
rs12909575	*ANXA2*	Intron	None	42/291/489	65/349/458	0.0076	0.692	0.895	37/241/298	29/148/177	0.58	0.803	0.023
rs4658963	*DISC1*	Intron	None	75/329/418	47/344/481	0.0077	0.692	0.150	39/227/303	22/151/178	0.64	0.176	0.036
rs4930387	*SPTBN2*	Downstream	None	162/445/215	189/409/274	0.0099	0.692	0.118	132/312/133	84/181/90	0.62	0.706	0.0049
rs4548	*RAB7*	Exon	Synonym change	0/96/726	3/75/794	0.020	0.868	0.392	0/62/510	2/42/311	0.20	0.655	0.015
rs10992429	*BICD2*	3' UTR	3' UTR	7/123/688	0/134/737	0.020	0.868	0.014	6/103/461	6/75/272	0.31	0.753	0.38
rs1545539	*ANXA2*	Intron	Regulatory (predicted)	24/211/587	31/270/571	0.032	0.921	0.895	10/184/387	14/95/249	0.042	0.202	0.10
rs7388	*RSN*	3′ UTR	3′ UTR	70/356/396	59/340/473	0.035	0.921	0.842	31/214/335	23/140/195	0.55	0.750	0.53
rs776746	*CYP3A5*	Intron	Splicing defect	8/104/710	1/103/768	0.039	0.921	0.196	2/66/508	0/38/318	0.65	0.287	0.029
rs7167571	*ANXA2*	Intron	None	55/273/494	40/328/504	0.049	0.930	0.144	35/205/333	19/121/212	0.81	0.750	0.12

Table 3 SNPs with nominal *P*-values <0.01 in the discovery sample excluded from replication sample

SNP	Gene	Location in gene or distance from the nearest gene (bp)	Genotype counts in British patients	Genotype counts in British controls	*P*-value in British	Reason for exclusion from replication set
rs6904085	Genomic control	317 034	396/329/97	384/421/67	0.00042	Genomic control
rs11862377	*DYNCILI2*	Intron	718/94/7	739/131/0	0.0014	r^2 0.75 with genotyped SNP rs9926649
rs363180	*DYNCILI2*	Upstream	761/56/5	783/88/0	0.0023	r^2 0.81 with genotyped SNP rs9926649
rs206695	Genomic control	500 075	478/310/34	453/361/58	0.0093	Genomic control

and 361 controls. Since the disease inheritance model is unknown, we chose SNPs for replication according to the genotypic test. We did not detect association with any of the genotyped SNPs in the replication sample (Table 2), further supporting the lack of association with any of the SNPs in the discovery sample.

Power calculations

We calculated the power to place an associated variant in the top 15 SNPs of the first stage of this study, and thus present as a candidate for our replication study and to replicate this variant in the second stage (Zaykin and Zhivotovsky, 2005). We assumed a multiplicative model with a causal allele frequency of 0.2 (mean observed minor allele frequency in replication set SNPs) and r^2 between the most associated marker allele and the causal variant of 0.865 (mean r^2 in our study). We found that a causal variant with an odds ratio of 1.47 had 80% power to both appear in the pool of top 15 SNPs in the first stage and be replicated in the second stage. We can therefore exclude well-tagged variants with effect sizes of OR > 1.47 in investigated genes as being associated with susceptibility to ALS, but we also note that due to some failure of genotyping it is not possible to guarantee that some causal variants may exist that have high effects sizes but low r^2 with any one SNP in our panel.

Discussion

To our knowledge this is the first large systematic pathway-based study to look at the role of common variation in susceptibility to sporadic ALS. Many lines of evidence from analysis of both human disease and animal models have implicated axonal transport defects in motor neuron death. Here we show that it is highly unlikely that common variation in the retrograde axonal transport and vesicle trafficking genes we assessed has a major effect on disease susceptibility, at least in the British population, as we did not find evidence of association in a large cohort of sporadic ALS patients and controls. Furthermore, we have excluded variants of moderate effect size by analysing the replication sample. Similarly, we did not find association with any of the genes in this study implicated in the metabolism of xenobiotic substances, with the focus on pesticide metabolism. A few of the genes—cytoplasmic dynein 1 heavy chain 1 (*DYNC1H1*) gene (Shah *et al.*, 2006), *SOD1* (Broom *et al.*, 2004), paraoxonase (*PON*) cluster (Saeed *et al.*, 2006; Slowik *et al.*, 2006)—included in our study have been investigated in small studies before and an association between ALS and PON genes has been suggested (Saeed *et al.*, 2006; Slowik *et al.*, 2006). We did not detect a signal of association with *PON1* and *PON2* genes in our large dataset; however, we note that we genotyped a different set of SNPs therefore the data are not directly comparable.

Given that such mechanisms, especially alterations in axonal transport, are generally thought to play a major role in motor neuron disease, the negative findings of our study are somewhat unexpected. We note, however, that we cannot exclude the relevance of these pathways to other neurodegenerative diseases such as spinal muscular atrophies. The only known familial human mutation in the dynein–dynactin complex, namely in *DCTN1* gene, results in an atypical form of slowly progressive lower motor neuron syndrome with major vocal cord involvement (Puls *et al.*, 2005), and mice mutations in *Dync1h1* give rise to a lower motor neuron phenotype similar to spinal muscular atrophies rather than ALS. Also, the human familial mutations in β-III spectrin (*SPTBN2*), a dynactin binding protein which is thought to serve as a link between cargoes and motor proteins, results in spinocerebellar ataxia type 5 (Ikeda *et al.*, 2006).

Furthermore, we cannot exclude the existence of rare causal or protective variants which were not assessed by our association approach. Multiple weakly deleterious variants can contribute substantially to susceptibility to complex disease, but they may not reach appreciable frequencies in populations, especially if they are under weak negative selection. In this regard we note that genome-wide analysis of patterns of diversity and divergence of protein-coding genes showed an excess of genes under weak negative selection among general vesicle transport and microtubule-binding motor protein genes (Bustamante *et al.*, 2005), and, for example four of nine dynein–dynactin complex subunit coding genes were found to be under weak negative selection in that study (comparing with 13.5% loci genome-wide). Furthermore, the dynein–dynactin complex

is involved in many essential cellular processes besides its role in axonal transport and a number of genes, including all dynein light chain genes are highly conserved among species (Pfister *et al*., 2006) and have very few common polymorphisms in coding sequences. All this indirectly supports involvement of these genes in complex diseases, but a multiple rare deleterious variants model may be more appropriate. Therefore an assessment of all variation in these genes in ALS and in other forms of motor neuron disease is needed to fully elucidate their possible role in motor neuron death.

The general weaknesses of many association studies in ALS as well as other diseases are inconclusive results, usually due to inability to discriminate among false positive findings, limited power to detect variants of moderate effect size, inconsistent replication results or no attempt at replication and investigating very small regions of the genome. We addressed all these issues in our study design, which therefore allows us to make firm conclusions about the investigated genes. However, our study had some limitations too. Selection of candidate genes relied on existing knowledge of the pathogenesis of ALS. We did not study all candidate genes; instead we chose a novel pathway approach and comprehensively screened genes of the dynein–dynactin complex and associated pathways, vesicle-trafficking genes and key enzymes involved in pesticide and other xenobiotics metabolism. Even so the choice of genes depended on existing biological knowledge of these pathways and we cannot completely exclude these pathways as there are certainly more genes directly or indirectly affecting retrograde axonal transport, vesicle trafficking and xenobiotic metabolism. In addition, not all common variants were sampled; however, the tagging approach let us cover most of the variation. The large discovery sample size enabled us to look for variants of moderate effect size and the replication sample allowed us to discriminate false positive findings. It may be that by genotyping only a subset of SNPs in the replication sample we missed variants with lower effect size and functional variants which were worse represented by the initially genotyped set of SNPs. Also, some variants may have population-specific effects, and this is illustrated by the association between the *VEGF* gene promoter polymorphisms and ALS (which is one of the strongest associations with ALS so far), which has been positive in three studies (Sweden, Belgium, Birmingham) and negative in another four (London, Sheffield, The Netherlands, North America) (Lambrechts *et al*., 2003; Van Vught *et al*., 2005; Brockington *et al*., 2005; Chen *et al*., 2006).

There are a few confounding factors which could have affected the results of our study, and these are relevant to all disease association studies. For example, patients were collected and phenotyped in different clinics over a lengthy period of time and we assume a uniformity of diagnosis among clinics. Different phenotype data collected in different clinics makes it difficult to compare the data and search for genes modifying disease progression and severity, and there is a degree of variation in the information available such as ethnicity or family history of the patient. Sample mix-up diminishes the power of the study, for example we detected 1–2% sex mismatches among the samples regardless of which clinic submitted the sample, indicating that the real magnitude of phenotyping errors is higher than this. From our experience in other association studies and anecdotal discussions with other laboratories carrying out large-scale human disease genetic association studies (J. Hardy, personal communication), we found that the number of errors in our study is within the range of that encountered in many similar scale projects. This points out the need for stringent quality control for handling samples in research laboratories and should be taken into account while creating DNA resources for future studies. However, we note that by genotyping large numbers of SNPs we were able to reduce the number of errors by detecting blindly duplicated samples and related individuals.

We detected a weak signal of differences in population structure between patient and control samples, even though most of the controls were collected in the same geographic areas as the cases. This is not likely to affect results and conclusions of our study; however, it must be taken into account planning future studies, especially on the same patient collections.

A further outcome of this study is that we validated the use of whole genome amplified DNA on a large scale, which may significantly ameliorate the use of existing DNA resources. An important factor limiting genetic studies of sporadic ALS until now was the lack of large sample collections providing enough power for association analysis and the lack of replication and attempts to replicate published findings. Due to low disease prevalence and short survival time, it takes many years and much effort to collect these irreplaceable samples and thus researchers are reluctant to 'waste' DNA to replicate marginally significant findings with little evidence of true positives. We show that about a 300-fold increase of DNA amount by whole genome amplification does not decrease genotyping quality and the decrease in genotyping efficiency is negligible (~1%).

In conclusion, we did not find evidence for association between retrograde axonal transport, vesicle trafficking and pesticide metabolism genes with ALS. However, we looked only at a small number of 'good' candidates, which we chose from the literature and experimental findings on ALS. Further large integrated well-designed studies, especially genome-wide screens, which are independent from prior knowledge on disease biology, are needed to elucidate genetic risk factors for sporadic ALS.

Supplementary material

Supplementary data are available at *Brain* online.

Acknowledgements

We thank Prof Dr Markus Boeck, Department of Transfusion Medicine and Immunohematology, University of Wuerzburg and Dr Simon Mead, MRC Prion Unit for control DNA samples; Dr Patrick F. Sullivan, University of North Carolina at Chapel Hill for analysing genes with TAMAL software. The study was funded by Medical Research Council Cooperative Group Component Grant G0400149, MND Association UK (KEM, HSP), Cancer Research UK (GS) and the MRC (DK, EMCF, EFG). The work of the Wuerzburg group was funded through the Deutsche Forschungsgemeinschaft through SFB 581 and EU funds through APOPIS. Funding to pay the Open Access publication charges for this article was provided by Medical Research Council Cooperative Group Component Grant G0400149.

References

Al-Chalabi A, Andersen PM, Nilsson P, Chioza B, Andersson JL, Russ C, et al. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. Hum Mol Genet 1999; 8: 157–64.

Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. Nature 2005; 437: 1299–320.

Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a practical and powerful approach to multiple testing. J Roy Stat Soc Ser B-Methodological 1995; 57: 289–300.

Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat 2001; 29: 1165–88.

Brockington A, Kirby J, Eggitt D, Schofield E, Morris C, Lewis CE, et al. Screening of the regulatory and coding regions of vascular endothelial growth factor in amyotrophic lateral sclerosis. Neurogenetics 2005; 6: 101–4.

Broom WJ, Parton MJ, Vance CA, Russ C, Andersen PM, Hansen V, et al. No association of the SOD1 locus and disease susceptibility or phenotype in sporadic ALS. Neurology 2004; 63: 2419–22.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, et al. Natural selection on protein-coding genes in the human genome. Nature 2005; 437: 1153–7.

Chance PF, Rabin BA, Ryan SG, Ding Y, Scavina M, Crain B, et al. Linkage of the gene for an autosomal dominant form of juvenile amyotrophic lateral sclerosis to chromosome 9q34. Am J Hum Genet 1998; 62: 633–40.

Chen W, Saeed M, Mao H, Siddique N, Dellefave L, Hung WY, et al. Lack of association of VEGF promoter polymorphisms with sporadic ALS. Neurology 2006; 67: 508–10.

Chen YZ, Bennett CL, Huynh HM, Blair IP, Puls I, Irobi J, et al. DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). Am J Hum Genet 2004; 74: 1128–35.

Corcia P, Camu W, Halimi JM, Vourc'h P, Antar C, Vedrine S, et al. SMN1 gene, but not SMN2, is a risk factor for sporadic ALS. Neurology 2006; 67: 1147–50.

Corcia P, Mayeux-Portas V, Khoris J, de Toffol B, Autret A, Muh JP, et al. Abnormal SMN1 gene copy number is a susceptibility factor for amyotrophic lateral sclerosis. Ann Neurol 2002; 51: 243–6.

de Bakker PIW, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005; 37: 1217–23.

de la Grange P, Dutertre M, Martin N, Auboeuf D. FAST DB: a website resource for the study of the expression regulation of human gene products. Nucleic Acids Res 2005; 33: 4276–84.

Devlin B, Bacanu SA, Roeder K. Genomic control to the extreme. Nat Genet 2004; 36: 1129–30.

Duncan JE, Goldstein LS. The genetics of axonal transport and axonal transport disorders. PLoS Genet 2006; 2.

Hadano S, Hand CK, Osuga H, Yanagisawa Y, Otomo A, Devon RS, et al. A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. Nat Genet 2001; 29: 166–73.

Hafezparast M, Klocke R, Ruhrberg C, Marquardt A, Ahmad-Annuar A, Bowen S, et al. Mutations in dynein link motor neuron degeneration to defects in retrograde transport. Science 2003; 300: 808–12.

Hemminger BM, Saelim B, Sullivan PF. TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits. Bioinformatics 2006; 22: 626–7.

Ikeda Y, Dick KA, Weatherspoon MR, Gincel D, Armbrust KR, Dalton JC, et al. Spectrin mutations cause spinocerebellar ataxia type 5. Nat Genet 2006; 38: 184–90.

Jackson M, Al-Chalabi A, Enayat ZE, Chioza B, Leigh PN, Morrison KE. Copper/zinc superoxide dismutase 1 and sporadic amyotrophic lateral sclerosis: analysis of 155 cases and identification of a novel insertion mutation. Ann Neurol 1997; 42: 803–7.

James PA, Talbot K. The molecular genetics of non-ALS motor neuron diseases. Biochim Biophys Acta 2006; 1762: 986–1000.

Jones CT, Shaw PJ, Chari G, Brock DJ. Identification of a novel exon 4 SOD1 mutation in a sporadic amyotrophic lateral sclerosis patient. Mol Cell Probes 1994; 8: 329–30.

Jones CT, Swingler RJ, Simpson SA, Brock DJ. Superoxide dismutase mutations in an unselected cohort of Scottish amyotrophic lateral sclerosis patients. J Med Genet 1995; 32: 290–2.

Kieran D, Hafezparast M, Bohnert S, Dick JRT, Martin J, Schiavo G, et al. A mutation in dynein rescues axonal transport defects and extends the life span of ALS mice. J Cell Biol 2005; 169: 561–7.

Lambrechts D, Storkebaum E, Morimoto M, Del Favero J, Desmet F, Marklund SL, et al. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. Nat Genet 2003; 34: 383–94.

LaMonte BH, Wallace KE, Holloway BA, Shelly SS, Ascano J, Tokito M, et al. Disruption of dynein/dynactin inhibits axonal transport in motor neurons causing late-onset progressive degeneration. Neuron 2002; 34: 715–27.

Ligon LA, LaMonte BH, Wallace KE, Weber N, Kalb RG, Holzbaur ELF. Mutant superoxide dismutase disrupts cytoplasmic dynein in motor neurons. Neuroreport 2005; 16: 533–6.

Mehta CR, Patel NR. Fexact - a Fortran subroutine for Fisher exact test on unordered Rxc contingency-tables. Acm Trans Math Softw 1986; 12: 154–61.

Munch C, Sedlmeier R, Meyer T, Homberg V, Sperfeld AD, Kurt A, et al. Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS. Neurology 2004; 63: 724–6.

Murakami T, Nagano I, Hayashi T, Manabe Y, Shoji M, Setoguchi Y, et al. Impaired retrograde axonal transport of adenovirus-mediated E-coli LacZ gene in the mice carrying mutant SOD1 gene. Neurosci Lett 2001; 308: 149–52.

Nelson LM, McGuire V. Epidemiology of amyotrophic lateral sclerosis. In: Brown RH, Swash M, Pasinelli P, editors. Amyotrophic lateral sclerosis. 2nd edn. 2006. p. 25–41.

Nishimura AL, Mitne-Neto M, Silva HC, Richieri-Costa A, Middleton S, Cascio D, et al. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. Am J Hum Genet 2004; 75: 822–31.

Parkinson N, Ince PG, Smith MO, Highley R, Skibinski G, Andersen PM, et al. ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B). Neurology 2006; 67: 1074–7.

Pasinelli P, Brown RH. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat Rev Neurosci 2006; 7: 710–23.

Pfister KK, Shah PR, Hummerich H, Russ A, Cotton J, Annuar AA, et al. Genetic analysis of the cytoplasmic dynein subunit families. PLoS Genet 2006; 2: e1.

Puls I, Jonnakuty C, LaMonte BH, Holzbaur ELF, Tokito M, Mann E, et al. Mutant dynactin in motor neuron disease. Nat Genet 2003; 33: 455–6.

Puls I, Oh SJ, Sumner CJ, Wallace KE, Floeter MK, Mann EA, et al. Distal spinal and bulbar muscular atrophy caused by dynactin mutation. Ann Neurol 2005; 57: 687–94.

Rosen DR. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature 1993; 364: 362.

Saeed M, Siddique N, Hung WY, Usacheva E, Liu E, Sufit RL, et al. Paraoxonase cluster polymorphisms are associated with sporadic ALS. Neurology 2006; 67: 771–6.

Schmitt-John T, Drepper C, Mussmann A, Hahn P, Kuhlmann M, Thiel C, et al. Mutation of Vps54 causes motor neuron disease and defective spermiogenesis in the wobbler mouse. Nat Genet 2005; 37: 1213–5.

Shah PR, Ahmad-Annuar A, Ahmadi KR, Russ C, Sapp PC, Horvitz HR, et al. No association of DYNC1H1 with sporadic ALS in a case-control study of a northern European derived population: a tagging SNP approach. Amyotroph Lateral Scler 2006; 7: 46–56.

Simpson CL, Al-Chalabi A. Amyotrophic lateral sclerosis as a complex genetic disease. Biochim Biophys Acta 2006; 1762: 973–85.

Skibinski G, Parkinson NJ, Brown JM, Chakrabarti L, Lloyd SL, Hummerich H, et al. Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. Nat Genet 2005; 37: 806–8.

Slowik A, Tomik B, Wolkow PP, Partyka D, Turaj W, Malecki MT, et al. Paraoxonase gene polymorphisms and sporadic ALS. Neurology 2006; 67: 766–70.

Van Vught PW, Sutedja NA, Veldink JH, Koeleman BP, Groeneveld GJ, Wijmenga C, et al. Lack of association between VEGF polymorphisms and ALS in a Dutch population. Neurology 2005; 65: 1643–5.

Veldink JH, Kalmijn S, Van der Hout AH, Lemmink HH, Groeneveld GJ, Lummen C, et al. SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. Neurology 2005; 65: 820–5.

Yang Y, Hentati A, Deng HX, Dabbagh O, Sasaki T, Hirano M, et al. The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. Nat Genet 2001; 29: 160–5.

Zaykin DV, Zhivotovsky LA. Ranks of genuine associations in whole-genome scans. Genetics 2005; 171: 813–23.