Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study


Summary

Background Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease of motor neurons that results in progressive weakness and death from respiratory failure, commonly within about 3 years. Previous studies have shown association of a locus on chromosome 9p with ALS and linkage with ALS–frontotemporal dementia. We aimed to test whether this genomic region is also associated with ALS in an independent set of UK samples, and to identify risk factors associated with ALS in a further genome-wide association study that combined data from the independent analysis with those from other countries.

Methods We collected samples from patients with sporadic ALS from 20 UK hospitals and obtained UK control samples from the control groups of the Depression Case Control study, the Bipolar Affective Case Control Study, and the British 1958 birth cohort DNA collection. Genotyping of DNA in this independent analysis was done with Illumina HumanHap550 BeadChips. We then undertook a joint genome-wide analysis that combined data from the independent set with published data from the UK, USA, Netherlands, Ireland, Italy, France, Sweden, and Belgium. The threshold for significance was p=0·05 in the independent analysis, because we were interested in replicating a small number of previously reported associations, whereas the Bonferroni-corrected threshold for significance in the joint analysis was p=2·20×10⁻⁷.

Findings After quality control, samples were available from 599 patients and 4144 control individuals in the independent set. In this analysis, two single nucleotide polymorphisms in a locus on chromosome 9p21.2 were associated with ALS: rs3849942 (p=2·22×10⁻⁶; odds ratio [OR] 1·39, 95% CI 1·21–1·59) and rs2814707 (p=3·32×10⁻⁶; 1·22, 1·15–1·30). In the joint analysis, which included samples from 4312 patients with ALS and 8425 control individuals, rs3849942 (p=4·64×10⁻¹⁰; OR 1·22, 95% CI 1·15–1·30) and rs2814707 (p=4·72×10⁻¹⁰; 1·22, 1·15–1·30) were associated with ALS.

Interpretation We have found strong evidence of a genetic association of two single nucleotide polymorphisms on chromosome 9 with sporadic ALS, in line with findings from previous independent GWAS of ALS and linkage studies of ALS–frontotemporal dementia. Our findings together with these earlier findings suggest that genetic variation at this locus on chromosome 9 causes sporadic ALS and familial ALS–frontotemporal dementia. Resequencing studies and then functional analysis should be done to identify the defective gene.

Funding ALS Therapy Alliance, the Angel Fund, the Medical Research Council, the Motor Neurone Disease Association of Great Britain and Northern Ireland, the Wellcome Trust, and the National Institute for Health Research Dementias and Neurodegenerative Diseases Research Network (DeNDRoN).

Introduction Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease of motor neurons that results in progressive weakness and death from respiratory failure, commonly within about 3 years. About one in 300 people develops ALS but its prevalence is low—five per 100 000 people—because of its poor prognosis. About 5% of patients have a family history of ALS, and occasionally frontotemporal dementia. ALS and frontotemporal dementia can coexist in the same family or individual. Several studies have reported linkage of ALS–frontotemporal dementia with a locus on chromosome 9p, and specifically with a shared region that spans about 3·6 Mb.2,3 A genome-wide association study (GWAS) reported that sporadic ALS was associated with single nucleotide polymorphisms (SNPs) in the same region of chromosome 9p.4 Therefore, this locus is the focus of research into the genetic basis of familial ALS, the relation between ALS and frontotemporal dementia, and the mechanisms underlying sporadic ALS and frontotemporal dementia.
Previous GWAS in ALS have reported association results that have been replicated within the same study but rarely in independent studies.8 For example, variation in the ITPR2 gene was associated with ALS in a GWAS and was replicated within the study9 but has not been found by other investigators.9 Variation in DPP6 was reported to be associated with ALS and was replicated within the same study,10 and a second study found that the same SNPs were the most strongly associated, although they did not achieve genome-wide significance.11 A subsequent study failed to confirm the findings.12 Variation in the UNC13A gene was strongly associated with ALS and was replicated within the same study,13 but no successful replication of the association with this locus has been reported.

We therefore aimed to examine previously reported genetic associations for ALS in a GWAS in an independent set of UK samples. We also aimed to combine these data with existing genotypic data from seven other countries to generate a large GWAS in ALS.

Methods

Samples

Adults attending one of 20 UK hospitals and who had been diagnosed with ALS by two consultant neurologists were invited to participate in the UK National DNA Bank for Motor Neuron Disease Research (MND DNA bank), a national collection of ALS DNA samples derived from whole blood. Collection began in 2004, and enrolment is ongoing. Patients had no family history of ALS at the time of sampling, were of self-reported white European ancestry, and had onset of weakness on or after January, 2002. Patients provided ethically approved written consent, and patients were selected if more than 1% of their genotypic data was missing; if they had abnormal heterozygosity, poor cluster separation, or a sex assignment that conflicted with phenotypic data; if they shared more than 5% of alleles identical by descent with any other study member, or if they were of non-European ancestry. SNPs with minor allele frequency less than 1%, showing non-random missingness between genotypes or cases and controls, or showing departure from Hardy-Weinberg equilibrium ($p<1\times10^{-8}$) were excluded.

For the joint analysis, raw genotypes were obtained from public databases or directly from research groups and merged into a single file for analysis, with a covariate file listing the site of origin. Genotyping quality and control measures for the studies in the joint analysis have been published previously.1,3,10,17 All data were deposited at the ALS online genetics database18 and the European genome-phenome archive.

Statistical analysis

We did power analysis with the Genetic Power Calculator.6 For the independent analysis there was 80% power to detect a typical previously reported association with an odds ratio (OR) of 1.2 and minor allele frequency of 0.25 at $p=0.05$. We report power at $p=0.05$ in the independent analysis because, although we have done a GWAS, we are interested in replicating association at five specified loci (FGGY, DPP6, ITPR2, UNC13A, and chromosome 9p21.2). For the joint analysis, we investigated the possibility of new associations within the genome and therefore we corrected for multiple testing. We used Bonferroni correction, which assumes the tests of association at each of the 227475 SNPs are independent; the corrected threshold is therefore $2.20\times10^{-7}$ ($0.05/227475$), and there was $83.5\%$ power to detect a similar variant at the threshold of $p=2.20\times10^{-7}$. Our aim was to use all of the available samples.

For the joint analysis, population stratification was controlled for by principal component analysis in EIGENSTRAT.9 We used the Tracy-Widom distribution to calculate the number of significant principal components needed as covariates in subsequent analyses, using the twstsv program of the Eigenssoft package.24

We tested association of genotypes with ALS by logistic regression with 30 principal components as covariates in the study regression. For the independent analysis there was 80% power to detect a similar variant at the threshold of $p=2\times10^{-7}$. Our aim was to use all of the available samples.

For the joint analysis, population stratification was controlled for by principal component analysis in EIGENSTRAT.9 We used the Tracy-Widom distribution to calculate the number of significant principal components needed as covariates in subsequent analyses, using the twstsv program of the Eigenssoft package.24

Genotyping of DNA from samples held with the UK MND DNA bank was done with Illumina HumanHap550 BeadChips (Illumina, CA, USA) by UCL Genomics, London, UK. All DNA samples went through stringent quality control, and processing was done under full laboratory information management system control. The raw data were analysed with GenomeStudio (Illumina) and extracted for statistical analysis.

Quality control procedures were applied to individual and SNP data (webappendix pp 1–3). Individuals were excluded if more than 1% of their genotypic data was missing; if they had abnormal heterozygosity, poor cluster separation, or a sex assignment that conflicted with phenotypic data; if they shared more than 5% of alleles identical by descent with any other study member, or if they were of non-European ancestry. SNPs with minor allele frequency less than 1%, showing non-random missingness between genotypes or cases and controls, or showing departure from Hardy-Weinberg equilibrium ($p<1\times10^{-8}$) were excluded.
was tested with a sliding window of varying sizes, including a window sufficient to include all SNPs in the haplotype block that contained the most strongly associated SNPs. We imputed genotypes by use of Impute2,23 with data from the 1000 Genomes Project and the International HapMap as templates.24 Genotype imputation allows the prediction of genotypes at untyped SNPs by the correlation between SNP genotypes in different populations. Association analysis can then be done on the predicted genotypes. Association analysis was done with SNPTEST,25 which accounts for uncertainty in the imputed genotypes, using the first ten principal components as covariates (the maximum possible). The correction for population stratification in the imputed data will have been weakened by the use of fewer principal components; however, the use of Impute2 increased our ability to deal with genotype uncertainty. We did this to identify variants for follow-up genotyping in a focused region, and thus the trade-off was deemed worthwhile.

Population attributable risk was estimated as:

\[
\frac{(p^2 + 2pqgAB) - 1}{p^2 + 2pqgAB + q^2gBB} \]

where \( q \) is the risk allele frequency, \( p = 1 - q \), and genotype relative risk is denoted by \( g \), with the subscript showing genotype. Penetration for each genotype was estimated as:

\[
f_{\text{case}} / f_{\text{control}}\]

where \( K \) is the lifetime prevalence and \( f \) is the genotype frequency.

Role of the funding source

The sponsor of the study had no role in study design, data analysis, data interpretation, writing of the report, or the decision to submit the paper for publication. The Motor Neurone Disease Association is the sponsor of the UK National MND DNA bank, and therefore had a role in data collection. The corresponding author had full access to all of the data in the study and had final responsibility for the decision to submit the paper for publication.

Results

For the independent analysis, genotypes were obtained from 663 individuals in the UK MND DNA bank, 1589 control samples from the DeCC study and BACCS, and 2930 control samples from the 1958 birth cohort. After quality control measures, samples were available from 599 people with ALS and 4144 control individuals. Table 1 summarises clinical features of the ALS cohort in the independent analysis.

For the joint analysis, genotypes were obtained from 4312 patients with ALS and 8425 control individuals. After quality control measures, samples were available from 4133 patients with ALS and 8130 control individuals, which were typed for 227475 SNPs to generate 2789513662 genotypes. Table 2 and the webappendix p 4 summarise demographic data.

In the independent analysis, SNPs rs3849942 (\( p = 2.22 \times 10^{-6} \); OR 1.39, 95% CI 1.21–1.59) and rs2814707 (\( p = 3.32 \times 10^{-6} \); 1.38, 1.20–1.58) on chromosome 9 showed association with ALS (figure 1, table 3, webappendix pp 5–6). rs903603, which had strong linkage disequilibrium with the associated SNPs (\( D' = 0.98 \)), was the SNP most strongly associated with ALS in the independent set (\( p = 8.92 \times 10^{-8}; \) OR 0.71, 95% CI 0.63–0.81). No other loci achieved genome-wide significance.

In the joint analysis, rs3849942 (\( p = 4.64 \times 10^{-10}; \) OR 1.22, 95% CI 1.15–1.30), rs2814707 (\( p = 4.72 \times 10^{-10}; \) 1.22, 1.15–1.30), and rs903603 (\( p = 1.37 \times 10^{-7}; \) 0.86, 0.82–0.91) showed association with ALS (figure 2, table 3, webappendix pp 7–8). SNPs at other loci that were associated with ALS in previous studies did not reach significance in the joint analysis, except the previously associated SNP in UNCI3A (rs12608932; table 3). The genotype relative risk for rs3849942 was 1.24 (heterozygote) and 1.49 (homozygote), with minor allele frequencies of 0.27 in patients with ALS and 0.24 in controls. The population attributable risk of rs3849942 was 0.10 and the penetrance of the risk allele was 0.004 (heterozygote) and 0.005 (homozygote).

A haplotypic association of rs3849942 and rs10122902 (omnibus \( p = 1.51 \times 10^{-11} \)) was identified in the joint analysis. Of the three haplotypes in the population (GA, AG, and GG), one had equal frequency in people with ALS and controls (GA: frequency cases 0.55, controls 0.52, controls 0.20; \( p = 0.123 \)), one was associated with increased risk of ALS (AG: frequency cases 0.27, controls 0.23, \( p = 7.76 \times 10^{-10} \)), and was one associated with reduced risk of ALS (GG: frequency cases 0.52, controls 0.57; \( p = 4.83 \times 10^{-11} \)). Similar results were obtained for the independent study: rs3849942 and rs10122902 showed

---

**Table 1: Clinical features of patients with amyotrophic lateral sclerosis in the independent analysis**

<table>
<thead>
<tr>
<th>Presentation</th>
<th>Case cohort (n=599)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulbar symptoms</td>
<td>165 (28%)</td>
</tr>
<tr>
<td>Limb weakness</td>
<td>375 (63%)</td>
</tr>
<tr>
<td>Respiratory</td>
<td>11 (2%)</td>
</tr>
<tr>
<td>Mixed†</td>
<td>42 (7%)</td>
</tr>
<tr>
<td>Age of onset (years)</td>
<td>60.7 (61.7, 22.9–86.5)</td>
</tr>
<tr>
<td>Diagnostic delay (months)</td>
<td>21.5 (18.3, 2.0–80.1)</td>
</tr>
</tbody>
</table>

Data are number (%) or mean (median, range). †Simultaneously reported weakness of limbs and bulbar musculature.
haplotypic association (omnibus \( p=1.01 \times 10^{-7} \)), and the same three haplotypes had similar control frequencies (GA: frequency cases 0.21, controls 0.19, \( p=0.074 \); AG: frequency cases 0.30, controls 0.24, \( p=3.18 \times 10^{-6} \); GG: frequency cases 0.48, controls 0.57, \( p=4.24 \times 10^{-8} \)). Longer SNP haplotypes did not result in stronger association in either part of the study (data not shown). The sliding window showed that pairs of SNPs within a 106.5 kb block of linkage disequilibrium flanked by rs4879515 and rs702231 were nearly all associated with ALS at \( p<1 \times 10^{-3} \), whereas outside this region there were no significant associations.

Imputation in the joint analysis in the region of interest (ie, the 3.6 Mb region identified by linkage) gave results for 21899 imputed SNPs that were derived from 519 typed SNPs. Five SNPs were more strongly associated with ALS than was rs3849942; the most strongly associated SNP was a rare variant, chr9:27573575 (minor allele frequency 0.054; \( p=6.40 \times 10^{-15} \)). The most strongly associated variant with a minor allele frequency greater than 0.1 was rs13295103 (\( p=1.97 \times 10^{-12} \); figure 3).

**Discussion**

In our independent study we have confirmed association of a chromosome 9 locus with ALS. We did not replicate any of the other previously reported associations, suggesting that our study gave a false-negative result or that they were population specific or false-positive results (panel, table 3). 8–12 In our joint analysis, which is the largest GWAS of ALS to date, chromosome 9p21.2 and

### Table 2: Characteristics of the component populations for the joint analysis after quality control

<table>
<thead>
<tr>
<th>Country</th>
<th>Cases</th>
<th>Controls</th>
<th>Total</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>286</td>
<td>307</td>
<td>593</td>
<td>313</td>
</tr>
<tr>
<td>France</td>
<td>230</td>
<td>700</td>
<td>930</td>
<td>805</td>
</tr>
<tr>
<td>Netherlands</td>
<td>977</td>
<td>988</td>
<td>1965</td>
<td>1152</td>
</tr>
<tr>
<td>Ireland</td>
<td>210</td>
<td>200</td>
<td>410</td>
<td>219</td>
</tr>
<tr>
<td>Italy</td>
<td>259</td>
<td>241</td>
<td>500</td>
<td>271</td>
</tr>
<tr>
<td>Sweden</td>
<td>422</td>
<td>441</td>
<td>863</td>
<td>481</td>
</tr>
<tr>
<td>USA</td>
<td>923</td>
<td>904</td>
<td>1827</td>
<td>1061</td>
</tr>
<tr>
<td>UK</td>
<td>227</td>
<td>205</td>
<td>432</td>
<td>264</td>
</tr>
<tr>
<td>UK MND DNA bank</td>
<td>599</td>
<td>0</td>
<td>599</td>
<td>382</td>
</tr>
<tr>
<td>DeCC study and BACCS</td>
<td>0</td>
<td>1505</td>
<td>1505</td>
<td>567</td>
</tr>
<tr>
<td>1958 birth cohort</td>
<td>0</td>
<td>2639</td>
<td>2639</td>
<td>1357</td>
</tr>
<tr>
<td>Total</td>
<td>4133</td>
<td>8130</td>
<td>12263</td>
<td>6872</td>
</tr>
</tbody>
</table>

The final three rows refer to samples used in the independent study.

DeCC = Depression Case Control. BACCS = Bipolar Affective Case Control Study.

**Figure 1:** p values for association in the independent genome-wide association study
Table 3: Associations for genes reported to be associated with amyotrophic lateral sclerosis in previous genome-wide association studies

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>n for associated SNP in previous studies*</th>
<th>Associated SNP</th>
<th>Controls</th>
<th>p value in previous studies</th>
<th>Independent analysis†</th>
<th>joint analysis‡</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>p value</td>
<td>Minor allele frequency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>FGGR</td>
<td>1287</td>
<td>rs6700125</td>
<td>1567</td>
<td>6.0×10⁻⁴</td>
<td>0.71</td>
<td>0.342</td>
</tr>
<tr>
<td>DPP6</td>
<td>1767</td>
<td>rs10260404</td>
<td>1916</td>
<td>5.4×10⁻⁸</td>
<td>0.89</td>
<td>0.397</td>
</tr>
<tr>
<td>ITPR2</td>
<td>1337</td>
<td>rs2306677</td>
<td>1356</td>
<td>3.28×10⁻⁸</td>
<td>0.98</td>
<td>0.100</td>
</tr>
<tr>
<td>LINC134A</td>
<td>4855</td>
<td>rs12608932</td>
<td>14953</td>
<td>2.53×10⁻¹⁰</td>
<td>0.42</td>
<td>0.366</td>
</tr>
<tr>
<td>9p21.21</td>
<td>4855</td>
<td>rs2814707</td>
<td>14953</td>
<td>7.45×10⁻⁸</td>
<td>3.32×10⁻⁴</td>
<td>0.305</td>
</tr>
<tr>
<td>9p21.21</td>
<td>4855</td>
<td>rs3849942</td>
<td>14953</td>
<td>1.01×10⁻⁸</td>
<td>2.22×10⁻⁸</td>
<td>0.304</td>
</tr>
<tr>
<td>9p21.21</td>
<td>4855</td>
<td>rs903603</td>
<td>14953</td>
<td>8.92×10⁻⁸</td>
<td>8.92×10⁻⁸</td>
<td>0.411</td>
</tr>
</tbody>
</table>

Previously reported significant associations and the results we obtained for those genes in the two parts of this study are shown.8–10 Total combined sample for the reported SNP, which includes a GWAS component and a non-GWAS follow-up study: 1599 cases and 4144 controls. 14133 cases and 8130 controls.

Figure 2: p values for association in the joint analysis
**Figure 3: Genetic architecture of the associated region**

Circles—log of the p value of association with amyotrophic lateral sclerosis for typed and imputed single nucleotide polymorphisms in the joint analysis, coloured according to linkage disequilibrium with rs3849942 (shown as a purple diamond) as measured by $r^2$. Blue lines—recombination rate across the locus based on the 1000 genomes project. Genes in the region are displayed below the graph. Arrows indicate the direction of transcription.

**UNC13A** were the only significantly associated loci. The association of **UNC13A** with ALS was weaker in the joint analysis than in the study that identified the association,\(^4\) which was well powered and replicated the results within the same study. Our independent samples do not show the association, possibly because the **UNC13A** variation is not a major cause of ALS in the UK, and so adding in these samples to the joint analysis we inevitably weaken the previously reported association. Given that the heritability of sporadic ALS is 0·61 (95% CI 0·38–0·78),\(^6\) our findings suggest that even larger GWASs are required to have sufficient power to detect genetic variations responsible for ALS. A potential limitation of this study is that not all control individuals were age and sex matched, although they were matched for geographical origin and ancestry. For a disease as rare as ALS, age and sex matching does not increase statistical power, whereas it would for a more common disorder.

The associated 9p21 locus is the focus of much interest for researchers of both ALS and frontotemporal dementia. Although frontotemporal dementia and ALS are independent diseases, there is some overlap. About 5% of people with ALS have overt frontotemporal dementia and up to 51% have subtle cognitive deficits that suggest frontal and temporal lobe dysfunction.\(^5\) Conversely, about 50% of people with frontotemporal dementia have features of motor neuron degeneration.\(^6\)

A locus on chromosome 9 has been independently linked to ALS–frontotemporal dementia in seven families;\(^7,8\) this 3·6 Mb locus is defined across studies by the flanking markers D9S169 and D9S251. The SNPs we have identified lie within this region, with the peak association at 106·5 Kb. A GWAS that used pathological subtyping of patients with frontotemporal dementia to increase homogeneity also identified the same SNPs as those most strongly associated with sporadic
frontotemporal dementia (rs3849942 $p=1\times10^{-6}$ and rs2814707 $p=5.72\times10^{-4}$). Therefore, variation in the same gene is likely to be responsible for sporadic ALS, sporadic frontotemporal dementia, and familial ALS–frontotemporal dementia. Assuming that the association locus is the same as the linkage locus, the region of interest in families with ALS–frontotemporal dementia must now be thought to be much narrower than that defined by recombination; however, if a synthetic association of rare variants and ALS is being detected, the disease-causing variation could still lie a substantial distance from the association peak.8

Penetrance is about 20% in the family with the strongest association between chromosome 9 and ALS–frontotemporal dementia.1 Thus, familial variants might underlie at least some of the sporadic cases of ALS and frontotemporal dementia. The large signal seen in our independent study of UK samples is consistent with a founder effect in the UK population, which possibly extends out to other populations. In this respect, that a GWAS of a cohort of Finnish patients with ALS and controls also detected the chromosome 9p locus (rs3849942 $p=9.1\times10^{-11}$) is interesting.11 This association signal was driven mainly by people with familial ALS; the most strongly associated SNP in this subset of patients was rs2225389 ($p=2.23\times10^{-12}$). All the patients with familial ALS who were not homozygous for the D90A allele of the SOD1 gene, and almost 20% of the patients with sporadic ALS, shared a 42-SNP haplotype, which strongly suggests a founder effect for the chromosome 9p locus, at least in Finland. However, the penetrance of the associated risk allele of rs3849942 in sporadic ALS in our study is only slightly higher than the background risk of ALS (0.36–0.45%), whereas the penetrance of the familial variants is about 20%; thus, the likelihood that the same variation is responsible for both familial and sporadic ALS is questionable, which makes a founder effect less likely. The significant association of rs3849942 with sporadic ALS in our joint analysis including multiple populations suggests that a common genetic variation might be contributing to disease risk. The two-SNP haplotype that was most strongly associated with ALS in the joint analysis (rs3849942-rs10122902) mimics the behaviour of the APOE association with Alzheimer’s disease; the three APOE alleles—ε2, ε3, and ε4—are in fact pairs of alleles at two SNPs forming a haplotype. When a haplotype is formed from alleles at two SNPs there are four possible haplotypes, although not all will necessarily exist in the population. For both the rs3849942-rs10122902 haplotype in ALS and the APOE haplotype in Alzheimer’s disease, only three of the four possible haplotypes are present, and the fourth does not exist in the population. In both cases, all identified haplotypes are common, and one is associated with risk of the disease (APOE ε4 in Alzheimer’s disease; AG in ALS), one is associated with protection from the disease (APOE ε2 in Alzheimer’s disease; GG in ALS), and one is neutral (APOE ε3 in Alzheimer’s disease; GA in ALS). Thus, common variation, as with APOE in Alzheimer’s disease, rather than a founder effect, could be responsible for sporadic ALS. However, because the penetrance at a functional variant might be much higher than at an associated tag SNP, which would account for the discrepancy between the penetrance for familial ALS and sporadic ALS, a founder effect cannot be excluded. There must be multiple variants worldwide because families with chromosome 9p-linked ALS–frontotemporal dementia do not all share a common haplotype, and so multiple founders must exist. However, the chromosome 9p locus is important in ALS because the population attributable risk estimate shows that it is involved in 10% of all cases of sporadic ALS. The disease-associated SNPs we have identified lie between two genes, C9orf72 and MOBKL2B, and are close to IFNK, which lies in an intron of MOBKL2B. The protein encoded by MOBKL2B is Mps one binder kinase activator-like 2B, which is involved in kinase regulation. On the opposite strand at the same location is IFNK, which encodes interferon κ precursor, which is important for immunity to viral infection. The SNPs comprising the most strongly associated haplotype (rs3849942 and rs10122902) and imputed SNP (chr9:27573575) overlap or lie within C9orf72, which encodes a protein that might be involved in cell development and possibly spermatogenesis.8

Panel: Research in context

Systematic review

To identify previously published genome-wide association studies (GWAS) in amyotrophic lateral sclerosis (ALS), we searched PubMed and the ALSOD database (http://alsod.iop.kcl.ac.uk), and contacted known ALS genetics research groups. Four GWAS reported findings that reached genome-wide significance;8–12 we designed the independent study to replicate these. Association studies were selected for the joint analysis if research diagnostic criteria for ALS were used, if microarrays of at least 300 000 single nucleotide polymorphisms were found, and if individual rather than pooled genotypes were available. Because the main finding of the independent and joint analyses was of an association to a region on chromosome 9 also known to be associated with familial ALS–frontotemporal dementia and sporadic frontotemporal dementia, we searched PubMed and contacted research groups and identified a further seven families with ALS–frontotemporal dementia12–14 and a GWAS of frontotemporal dementia.15 These linkage data have been well summarised in a linkage paper.4

Interpretation

The ALS-associated locus on chromosome 9p was independently replicated, and no other loci were identified, suggesting that even larger genome-wide association studies are needed. We have shown the association signal on chromosome 9p lies in a well circumscribed block of linkage disequilibrium of about 106·5 Kb in size, which is flanked by markers rs4879515 and rs702231 and includes three annotated genes. This locus has also previously been implicated in sporadic frontotemporal dementia and in families with ALS–frontotemporal dementia. Because there is pathological and clinical overlap of ALS and frontotemporal dementia, the most logical interpretation is that dysfunction of a single gene product is responsible for some cases of ALS, frontotemporal dementia, and ALS–frontotemporal dementia.
Although linkage of ALS–frontotemporal dementia to this region of chromosome 9 was first identified in 2006,2,3 sequencing has not yet yielded a disease-causing variant such as a non-synonymous SNP or splice-site variation. If this remains the case, one explanation could be that the variation is tagged by SNPs but not easily detected by exome capture and high throughput sequencing or Sanger sequencing. The disease-causing variant might be in an unknown exon or gene and has therefore not been sequenced, or it might be a trinucleotide repeat expansion in an intron, although such repeats are not seen in the surrounding region. Copy number variation is not necessarily easily detected, depending on the sequencing method used, but three microarray-based studies in ALS have not detected ALS-associated copy number variation in this region of chromosome 9.4–6 Finally, a microRNA or non-annotated gene that has not been examined, or an inversion, which would be difficult to detect by sequencing, might be the disease-causing variant. Our results suggest genetic variation at this locus on chromosome 9 causes sporadic amyotrophic lateral sclerosis and familial ALS–frontotemporal dementia. Resequencing studies and then functional analyses are needed to identify the defective gene.

Acknowledgments
The independent study case samples were obtained from the UK DNA bank for Motor Neuron Disease Research, which was funded by the Motor Neurone Disease Association grant 3/3 and the Wellcome Trust grant 070222/A/02/Z, with additional support from the National Institute for Health Research Dementias and Neurodegenerative Diseases Research Network (DeNDRoN). Genotyping of the independent study case samples was funded by grants from the Angel Fund and ALS Therapy Alliance, with additional funding to RO and JH from the Motor Neurone Disease Association for genotyping (grant 10/57). We acknowledge use of genotyping data from the British 1958 birth cohort DNA collection, which was funded by the Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/2/02. Controls for the DeCC study and BACCS were genotyped under the Medical Research Council grant G0701420. This project has been supported by the Princes Beatrix Fonds, VSB Fonds, The Kersten Foundation, The Netherlands ALS Foundation, and The Adessium Foundation (LHvdB). JHV is supported by grants from The Brain Foundation of the Netherlands. WR is supported by grants from the University of Leuven (Methusalem) and the Interuniversity Attraction Poles (IUAP) program P6/43 of the Belgian Federal Science Policy Office. VS is supported by the Italian Ministry of Health (Ricerca Finalizzata 2007 number 31). This work was supported in part by the Intramural Research Program of the National Institutes of Health, and the National Institute on Aging (Z01-AG000949-02). AC is supported by the Italian Ministry of Health (Ricerca Sanitaria Finalizzata, 2007, number 30), the Fondazione Viali e Mauro per la SLA (grant 3/2008), and the Federazione Italiana Gioco Calcio (Grant 2/2008). We thank the ALS Association, the ALS Society of Canada, MND Iceland, and the National Institute for Health Research Biomedical Research Centre for Mental Health for support, and Kuang Lin for programming advice.

References


19 Purcell S, Cherny SS, Sham PC. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003; 19: 149–50.


