A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis


1Department of Neuroscience, 2Department of Clinical Neuroscience and 3MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King’s College London, London SE5 8AF, UK, 4Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, 20149 Milano, Italy, 5Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center and 6Doctoral School in Molecular Medicine, Università degli Studi di Milano, 20122 Milano, Italy, 7Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, 20133 Milano, Italy, 8Department of Psychological Medicine and Neurology, School of Medicine, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff CF14 4XN, UK, 9Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases, ‘A. Avogadro’ University, 28100 Novara, Italy, 10Department of Neurosciences, University of Padova, 35128 Padova, Italy, 11Laboratory of Experimental Neurobiology and 12Division of General Neurology, IRCCS ‘C. Mondino’ National Institute of Neurology Foundation, 27100 Pavia, Italy, 13Neurologic Unit, IRCCS Foundation Ca’ Granda Ospedale Maggiore Policlinico, 20122 Milano, Italy, 14Laboratory of Molecular Biology, IRCCS Istituto Auxologico Italiano, 20095 Cusano Milanino, Italy, 15Department of Neurology, ALS Center, Maggiore della Carità University Hospital, 28100 Novara, Italy, 16Institute of Neurological Sciences, National Research Council, 87050 Mangone, Italy, 17Division of Neurology, Department of Clinical and Experimental Medicine, University of Pisa, Nuovo Ospedale S.Chiara-Cisanello, 56126 Pisa, Italy, 18Department of Neurology and Psychiatry, University of Rome ‘Sapienza’, 00185 Roma, Italy, 19Department of Neurological Sciences, University Federico II, 80131 Napoli, Italy, 20Neurorehabilitation Unit, Azienda Ospedaliero-Universitaria Pisana, 56126 Pisa, Italy, 21Department of Laboratory Medicine, Medical Genetics Unit, Niguarda Ca’ Granda Hospital, 20162 Milano, Italy, 22NeuroMuscular Omnicentre (NEMO), Fondazione Serena Onlus, Niguarda Ca’ Granda Hospital, 20162 Milano, Italy, 23Department of

*To whom correspondence should be addressed at: Department of Neuroscience, Institute of Psychiatry, King’s College London, De Crespigny Park, London SE5 8AF, UK. Tel: +44 207848 5244; Fax: +44 2077080017; Email: isabella.fogh@kcl.ac.uk
†These authors contributed equally to the study.
‡These two senior authors contributed equally to the study.
§See Appendix for Slagen Consortium members.

© Crown copyright 2013.
Identification of mutations at familial loci for amyotrophic lateral sclerosis (ALS) has provided novel insights into the aetiology of this rapidly progressing fatal neurodegenerative disease. However, genome-wide association studies (GWAS) of the more common (~90%) sporadic form have been less successful with the exception of the replicated locus at 9p21.2. To identify new loci associated with disease susceptibility, we have established the largest association study in ALS to date and undertaken a GWAS meta-analytical study combining 3959 newly genotyped Italian individuals (1982 cases and 1977 controls) collected by SLAGEN (Italian Consortium for the Genetics of ALS) together with samples from Netherlands, USA, UK, Sweden, Belgium, France, Ireland and Italy collected by ALSGEN (the International Consortium on Amyotrophic Lateral Sclerosis Genetics). We analysed a total of 13 225 individuals, 6100 cases and 7125 controls for almost 7 million single-nucleotide polymorphisms (SNPs). We identified a novel locus with genome-wide significance at 17q11.2 (rs34517613 with $P = 1.11 \times 10^{-8}$; OR 0.82) that was validated when combined with genotype data from a replication cohort ($P = 8.62 \times 10^{-9}$; OR 0.833) of 4656 individuals. Furthermore, we confirmed the previously reported association at 9p21.2 (rs3849943 with $P = 7.69 \times 10^{-8}$; OR 1.16). Finally, we estimated the contribution of common variation to heritability of sporadic ALS as ~12% using a linear mixed model accounting for all SNPs. Our results provide an insight into the genetic structure of sporadic ALS, confirming that common variation contributes to risk and that sufficiently powered studies can identify novel susceptibility loci.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a late-onset progressive neurodegenerative disorder mainly affecting motor neurones. ALS is the most common adult onset motor neurone disease with prevalence of 5 per 100 000 and with family history, age and male gender as the major risk factors (1–4). While familial ALS is well characterized with several causative genes identified to date, the genetic architecture of the more common sporadic form is poorly understood. Previous genome-wide association studies (GWAS) have identified several loci associated with ALS risk such as DPP6, ITPR2, FGGY and UNC13a (5–11) that have failed to be replicated in independent populations (12–15). One exception is a locus on chromosome 9p21 (15) that has been reliably replicated (16). At this locus, an expanded hexanucleotide repeat in the C9orf72 gene has been recently identified as the causative mutation in a large proportion of familial (23–47%) as well as sporadic (~5%) ALS cases (17–20). This finding indicates that well-powered studies can identify novel loci associated with ALS susceptibility. To discover further loci, we designed a large GWAS meta-analysis with sufficient power to detect risk alleles with small effect sizes as observed in other neurodegenerative and complex diseases. There is evidence for a strong genetic component in...
sporadic ALS with heritability estimated to be 0.61 (95% CI: 0.38–0.78) in a combined study of 171 ALS twin pairs collected by three independent studies from Britain and Sweden (21), so we sought to estimate heritability derived from unrelated individuals and explained by common variation using GWAS data. Phenotype variation in complex traits is owing to interactions of genetic and environmental factors; thus, the quantification of the genetic variance is relevant in the study of multifactorial diseases. Heritability, the proportion of phenotype explained by genetic variance, is typically estimated in closely related individuals such as twin pairs, but this can inflate estimates as a consequence of epistatic interactions or shared environment (22,23). In contrast, heritability explained by associated SNPs identified in GWAS as passing-accepted thresholds of significance (typically $P \leq 5 \times 10^{-8}$) explains only a small fraction of the genetic variation in most complex diseases. The difference between the phenotypic variance explained by GWAS results and those estimated in family studies is referred to as the ‘missing heritability problem’ and possibly is explained by incomplete linkage disequilibrium (LD) between genotyped SNPs and causal variants and/or by the presence of gene-by-gene or gene-by-environment interactions (22,23). We have estimated heritability of ALS considering all SNPs simultaneously regardless of their association with ALS phenotype and compared this with heritability from twin studies.

**RESULTS**

Association analyses

We analysed genotype raw data from eight independent studies including 3959 newly genotyped Italian individuals (1982 cases and 1977 controls) and 11 611 individuals (5195 cases and 6416 controls) from previously published studies. Full description of the sample size included in this study is reported in Table 1 and Supplementary Material, Table S1.

Raw data for each study were assessed for quality control (QC) separately following the same criteria (see Supplementary Material and Table S2 for full discussion). Population structure of the individual cohorts was studied by means of principal components analysis (PCA) with EIGENSTRAT software (24,25), and outliers were identified by the projection of the first 10 principal components (Supplementary Material, Methods and Fig. S1).

To achieve the maximal coverage, we imputed genome-wide filtered originally genotyped data for each study (see Materials and methods). Overall, the average number of inferred genotypes was 8 342 920 SNPs varying proportionally to the original genotyping platform. After filtering (Supplementary Material, Methods), the eight datasets included genotype data for almost 7 million SNPs in 13 225 individuals (6100 cases and 7125 controls) (Supplementary Material, Table S2). Cleaned imputed genotypes were tested for association with ALS separately using SNPTESTv2.4.0 (26), and the logistic regression analyses were adjusted by the appropriate principal component axes estimated in the individual cohorts (Supplementary Material, Table S3). After genomic inflation control, we combined the logistic regression analysis results of each study in a meta-analysis using the program METAL (27). We adopted the ‘standard error’ analysis scheme, which combines effect size estimates ($\beta$-coefficients) across the studies weighted according to the inverse of the corresponding standard errors. As a result of the meta-analysis, we observed two loci reaching genome-wide threshold of significance, rs3849943 with $P = 7.69 \times 10^{-9}$ (OR 1.16, 95% CI: 1.10–1.22; average posterior probability (APP) 0.999; minor allele frequency (MAF) cases 0.268, controls 0.238) on chromosome 9p21 and SNP rs34517613 with $P = 1.11 \times 10^{-6}$ (OR 0.82, 95% CI: 0.76–0.87; APP 0.9279; MAF cases 0.108, controls 0.129) on chromosome 17q11.2.

A third locus, rs1788776, at 18q11.2 was very close to genome-wide significance threshold ($P = 7.67 \times 10^{-8}$, OR 0.87, 95% CI: 0.76–0.87; APP 0.9716; MAF cases 0.392, controls 0.362) (Table 2, Figs. 1 and 2, Supplementary Material, Fig. S2). With the exception of the locus on 9p21, loci previously found to be associated with ALS risk (5–11) did not reach genome-wide significance in our meta-analysis (Supplementary Material, Table S4) although SNP rs12608932 (UNC13A) at 19p13.3 had not complete coverage across datasets as it failed the QC threshold in the British dataset (Supplementary Material, Fig. S3). Additionally, examination of these previously associated loci in the Italian cohort as an independent replication study only found evidence for association for the 9p21 locus, and for the ITPR2 locus while significant, direction of effect is the opposite to that reported in the original study (8) (Supplementary Material, Fig S3). Only the ITPR2 locus showed significant evidence for heterogeneity across the different cohorts of European ancestry analysed here (Supplementary Material, Table S4).

### Table 1. Sample size and genotype platforms

<table>
<thead>
<tr>
<th>Source</th>
<th>Sample ancestry</th>
<th>Sample size</th>
<th>Cases (N)</th>
<th>Control (N)</th>
<th>Genotyping Illumina arrays</th>
<th>SNPs before imputation</th>
<th>SNPs after imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLAGEN Consortium</td>
<td>Italian</td>
<td>3959</td>
<td>1982</td>
<td>1977</td>
<td>660K</td>
<td>657366</td>
<td>8391895</td>
</tr>
<tr>
<td>UMC Utrecht</td>
<td>Dutch</td>
<td>911</td>
<td>461</td>
<td>450</td>
<td>317K</td>
<td>317503</td>
<td>8328981</td>
</tr>
<tr>
<td>Utrecht, Umeå, Leuven</td>
<td>Dutch, Swedish, Belgian</td>
<td>2806</td>
<td>1364</td>
<td>1442</td>
<td>370K</td>
<td>317503</td>
<td>8372645</td>
</tr>
<tr>
<td>MGH, King’s College, Evry</td>
<td>Northern American, British, French</td>
<td>3916</td>
<td>1710</td>
<td>2206</td>
<td>370K</td>
<td>307790</td>
<td>8370626</td>
</tr>
<tr>
<td>Beaumont Hospital, Dublin</td>
<td>Irish</td>
<td>432</td>
<td>221</td>
<td>211</td>
<td>550K</td>
<td>561466</td>
<td>8268635</td>
</tr>
<tr>
<td>NIH publicly available, Coriell biobank</td>
<td>Northern American</td>
<td>547</td>
<td>276</td>
<td>271</td>
<td>550K</td>
<td>555352</td>
<td>8325506</td>
</tr>
<tr>
<td>MND DNA bank and RADIANt study</td>
<td>British</td>
<td>2252</td>
<td>663</td>
<td>1589</td>
<td>610K</td>
<td>471994</td>
<td>8372429</td>
</tr>
<tr>
<td>NIH publicly available</td>
<td>Italian</td>
<td>747</td>
<td>500</td>
<td>247</td>
<td>550K</td>
<td>5000002</td>
<td>8312642</td>
</tr>
</tbody>
</table>

Independent studies analysed in the combined international meta-analysis were genotyped on different type of Illumina platforms. Italian data were collected by the SLAGEN Consortium (Italy) and the Dutch, Swedish, French, Belgian, Irish and North American, and Italian raw data were collected by the International ALS-GWAS Consortium (ALSGEN). British cases were collected by UK National MND Bank samples and British controls by the RADIANt study (DeCC, BACCs) and NIH publicly available data from Coriell biobank (www.coriell.org).
### Table 2. ALS-GWAS meta-analysis results

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>SNP</th>
<th>Position</th>
<th>A1</th>
<th>A2</th>
<th>APP</th>
<th>Info</th>
<th>MAF_ALS</th>
<th>MAF_CONT</th>
<th>OR</th>
<th>P-value</th>
<th>Direction</th>
<th>Low 95% CI</th>
<th>Up 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>rs3849943</td>
<td>27533382</td>
<td>C</td>
<td>T</td>
<td>1</td>
<td>0.999</td>
<td>0.268</td>
<td>0.238</td>
<td>1.166</td>
<td>7.69E−09</td>
<td>+−+++++++</td>
<td>1.107</td>
<td>1.228</td>
</tr>
<tr>
<td>9</td>
<td>rs3849941</td>
<td>27528948</td>
<td>A</td>
<td>G</td>
<td>0.993</td>
<td>0.986</td>
<td>0.266</td>
<td>0.236</td>
<td>1.166</td>
<td>9.54E−09</td>
<td>−++++−−</td>
<td>1.106</td>
<td>1.229</td>
</tr>
<tr>
<td>9</td>
<td>rs2477523</td>
<td>27526894</td>
<td>T</td>
<td>A</td>
<td>0.997</td>
<td>0.992</td>
<td>0.269</td>
<td>0.239</td>
<td>1.163</td>
<td>1.06E−08</td>
<td>−−−−−−</td>
<td>1.104</td>
<td>1.225</td>
</tr>
<tr>
<td>9</td>
<td>rs2453555</td>
<td>27553868</td>
<td>G</td>
<td>A</td>
<td>0.995</td>
<td>0.988</td>
<td>0.27</td>
<td>0.239</td>
<td>1.163</td>
<td>1.07E−08</td>
<td>−−−−−−</td>
<td>1.104</td>
<td>1.225</td>
</tr>
<tr>
<td>17</td>
<td>rs3451761</td>
<td>23634379</td>
<td>C</td>
<td>T</td>
<td>0.922</td>
<td>0.734</td>
<td>0.108</td>
<td>0.129</td>
<td>0.822</td>
<td>1.11E−08</td>
<td>++−−−−−</td>
<td>0.769</td>
<td>0.879</td>
</tr>
<tr>
<td>17</td>
<td>rs2937920</td>
<td>23614784</td>
<td>C</td>
<td>T</td>
<td>0.951</td>
<td>0.854</td>
<td>0.125</td>
<td>0.149</td>
<td>0.83</td>
<td>1.24E−07</td>
<td>++++++++</td>
<td>0.775</td>
<td>0.889</td>
</tr>
<tr>
<td>17</td>
<td>rs900055</td>
<td>23620255</td>
<td>C</td>
<td>T</td>
<td>1</td>
<td>1</td>
<td>0.217</td>
<td>0.249</td>
<td>0.857</td>
<td>1.98E−07</td>
<td>++++++++</td>
<td>0.809</td>
<td>0.908</td>
</tr>
<tr>
<td>17</td>
<td>rs1696420</td>
<td>23606837</td>
<td>G</td>
<td>C</td>
<td>0.978</td>
<td>0.954</td>
<td>0.224</td>
<td>0.257</td>
<td>0.86</td>
<td>2.55E−07</td>
<td>−−−−−−</td>
<td>0.812</td>
<td>0.911</td>
</tr>
<tr>
<td>17</td>
<td>rs2292633</td>
<td>2359507</td>
<td>G</td>
<td>A</td>
<td>0.956</td>
<td>0.906</td>
<td>0.221</td>
<td>0.254</td>
<td>0.86</td>
<td>2.82E−07</td>
<td>+++−−−−</td>
<td>0.812</td>
<td>0.911</td>
</tr>
<tr>
<td>17</td>
<td>rs3571469</td>
<td>23743915</td>
<td>G</td>
<td>A</td>
<td>0.973</td>
<td>0.934</td>
<td>0.169</td>
<td>0.192</td>
<td>0.845</td>
<td>3.21E−07</td>
<td>+++++++++</td>
<td>0.792</td>
<td>0.901</td>
</tr>
<tr>
<td>17</td>
<td>rs739439</td>
<td>2374949</td>
<td>C</td>
<td>T</td>
<td>0.994</td>
<td>0.964</td>
<td>0.169</td>
<td>0.193</td>
<td>0.842</td>
<td>3.97E−07</td>
<td>++++++++</td>
<td>0.788</td>
<td>0.900</td>
</tr>
<tr>
<td>17</td>
<td>rs900311</td>
<td>23619554</td>
<td>G</td>
<td>A</td>
<td>0.997</td>
<td>0.992</td>
<td>0.219</td>
<td>0.252</td>
<td>0.858</td>
<td>4.58E−07</td>
<td>++++++++</td>
<td>0.808</td>
<td>0.911</td>
</tr>
<tr>
<td>18</td>
<td>rs178776</td>
<td>19490840</td>
<td>G</td>
<td>A</td>
<td>0.971</td>
<td>0.947</td>
<td>0.392</td>
<td>0.362</td>
<td>0.872</td>
<td>7.67E−08</td>
<td>+++++−−−−</td>
<td>0.830</td>
<td>0.917</td>
</tr>
<tr>
<td>18</td>
<td>rs162935</td>
<td>19475137</td>
<td>C</td>
<td>G</td>
<td>0.956</td>
<td>0.927</td>
<td>0.383</td>
<td>0.353</td>
<td>0.874</td>
<td>1.00E−07</td>
<td>−−−−−−</td>
<td>0.832</td>
<td>0.918</td>
</tr>
<tr>
<td>18</td>
<td>rs6209304</td>
<td>19489968</td>
<td>T</td>
<td>C</td>
<td>0.979</td>
<td>0.96</td>
<td>0.389</td>
<td>0.36</td>
<td>0.875</td>
<td>2.09E−07</td>
<td>−−−−−−</td>
<td>0.832</td>
<td>0.920</td>
</tr>
<tr>
<td>18</td>
<td>rs1295785</td>
<td>19487958</td>
<td>G</td>
<td>T</td>
<td>0.988</td>
<td>0.976</td>
<td>0.408</td>
<td>0.382</td>
<td>0.878</td>
<td>2.64E−07</td>
<td>++++++++</td>
<td>0.836</td>
<td>0.923</td>
</tr>
<tr>
<td>18</td>
<td>rs1652373</td>
<td>19482645</td>
<td>G</td>
<td>T</td>
<td>1</td>
<td>1</td>
<td>0.409</td>
<td>0.382</td>
<td>0.877</td>
<td>2.66E−07</td>
<td>++++++++</td>
<td>0.834</td>
<td>0.922</td>
</tr>
<tr>
<td>18</td>
<td>rs1206021</td>
<td>19499748</td>
<td>C</td>
<td>T</td>
<td>0.979</td>
<td>0.961</td>
<td>0.408</td>
<td>0.381</td>
<td>0.88</td>
<td>3.29E−07</td>
<td>+++++++</td>
<td>0.838</td>
<td>0.924</td>
</tr>
<tr>
<td>18</td>
<td>rs1711468</td>
<td>19498537</td>
<td>C</td>
<td>A</td>
<td>0.97</td>
<td>0.945</td>
<td>0.41</td>
<td>0.384</td>
<td>0.882</td>
<td>4.16E−07</td>
<td>++++++++</td>
<td>0.840</td>
<td>0.926</td>
</tr>
</tbody>
</table>

Summary of the most significant associated SNPs in the international ALS-GWAS meta-analysis from imputed genotypes. Meta-analysis (METAL) was performed weighting by the $β$-coefficient estimates and the inverse of their corresponding standard errors. Effect size of reference allele A1 is expressed with positive and negative symbols; symbols ‘+’ and ‘−’ indicate the direction of the effect size (beta values from regression), where a plus symbol means that increasing frequency of allele A1 is correlated with increasing trait values and vice versa.

Human Molecular Genetics, 2014, Vol. 23, No. 8
Further larger studies are needed to confirm the role of these loci in ALS susceptibility.

As inclusion of non-confounding covariates such as gender and age can significantly reduce power to identify truly associated variants particularly when the disease prevalence is low (<1%), we have not included non-confounding covariates in logistic regression analyses presented here (28). In confirmation of our approach, the most significant SNPs ($P \leq 1 \times 10^{-8}$) mapping in the two candidate loci at 17q and 18q showed no gender-by-SNP interaction in multi regression analysis (see Materials and methods).

For these loci, we searched for secondary signals by performing logistic regression conditioning upon the most associated SNPs under an additive model. There was no evidence of independent SNPs in any of these regions (Supplementary Material, Table S6). As the international meta-analysis included samples from Northern and Southern Europe, we tested the most associated SNPs for genetic heterogeneity quantifying the effect of the degree of variation between studies (Supplementary Material, Methods). We observed no significant heterogeneity in the distribution of allele frequencies between European populations for these SNPs (Supplementary Material, Table S7).

As additional evidence, we tested the candidate loci at 17q and 18q in a novel independent replication cohort of 2074 cases and 2556 controls collected in Italy, Netherlands and Germany (see full description in the Supplementary Material, Methods and Table S8). For each locus, we selected two variants, the most associated in the GWAS meta-analysis (imputed data) and

Figure 1. Manhattan plot of the international ALS meta-analysis. Scatter plot of chromosome position (x-axis) against $-\log_{10}$ GWAS meta-analysis $P$-values (y-axis) from imputed data. The threshold of genome-wide significance ($P = 5 \times 10^{-8}$) is indicated as a horizontal red line. At locus 9p21.2, 18 SNPs close to Corf72 gene lie above the red line (most significant SNP is rs3849943 with $P = 7.69 \times 10^{-9}$). Locus 17q11.2 shows SNP rs34517613 ($P = 1.11 \times 10^{-8}$) to be significantly associated. SNP rs1788776 at 18q11.2 is very close to the threshold of significance with $P = 7.67 \times 10^{-8}$. Manhattan plot was produced using ggplot2 package in R.

Figure 2. Regional association plots of the associated loci. LD structure of the three regions associated with ALS in the international GWAS meta-analysis. For each plot, the $-\log_{10} P$-values (left y-axis) of SNPs are shown according to their chromosomal positions (x-axis); the genetic recombination rates are shown on the right y-axis. $-\log_{10} P$-values are shown for both genotyped and imputed SNPs distributed in a 0.8-megabase genomic region. The top SNP of each region is indicated as a diamond, and SNPs colour reflects LD correlation ($r^2$). (A) locus at 9p21, (B) 17q11.2 and (C) 18q11.2. LD plots were generated by LocusZoom v1.1.
combined joint analysis (original genotype data) (Supplementary Material, Table S9). We found suggestive evidence for association in the independent replica analysis for SNP rs34517613 at 17q (P = 0.055; OR 0.89, 95% CI: 1.00–0.78) and when combined in meta-analysis with the full GWAS data, the association identified in the discovery sample became more significant (Pcombined = 8.62 × 10−5; OR 0.83, 95% CI: 0.891–0.778). We found no additional evidence for association at the 18q11.2 locus (Supplementary Material, Table S10).

The locus at 17q11.2 is gene rich with a number of plausible candidates (Fig. 2). In an effort to refine associations at this locus and at 18q11.2, we searched for non-synonymous SNPs in LD (r2 ≥ 0.4) with our two lead signals to identify possible functional variants that explained the associations with ALS risk. No non-synonymous variants were found in LD at the 17q11.2 locus; however, three SNPs, rs739439 (r2 = 0.416) in the 3′ untranslated region, the intronic SNPs rs35714695 (r2 = 0.457) and rs34660379 (r2 = 0.412) within the SARM1 (sterile alpha and TIR-containing motif 1) gene were in LD (Table 2, Fig. 2).

At the suggestive 18q11.2 locus, the lead SNP rs1788776 was only in strong LD with SNPs in the ANKRD29 gene (Fig. 2), an ankyrin repeat domain-containing protein of undetermined function, these included missense SNPs (rs12956232, r2 = 0.78; rs1788758, r2 = 0.78; rs12960692, r2 = 0.74; rs11662113, r2 = 0.53) (Table 2, Fig. 2).

Next, we looked for cis eQTL (expression quantitative trait loci) in LD with our lead SNPs. All SNPs in LD (r2 ≥ 0.2) with rs34517613 and rs1788776 were analysed for association with cis eQTL using several publicly available eQTL databases (Supplementary Material), but none reached the P-value threshold for significance. This should not be considered conclusive as expression data from central nervous system tissues are still rather limited across eQTL databases and further studies should be carried out.

**Contribution of C9orf72 to chromosome 9p21 association signal**

As chromosome 9p21 was found to be strongly associated in our meta-analysis, we further investigated the contribution of ALS sporadic cases that also carry the expanded hexanucleotide repeat in the C9orf72 gene to this association. Information about carriers of the pathologic expansion in C9orf72 gene was available for 2287 of 6100 (37.3%) ALS cases from the Italian, Dutch and British studies included in the meta-analysis. The frequency of the expanded repeats carriers in the sporadic cases we have analysed progressively decreased as you proceed from Northern to Southern Europe ranging from 8 to 4.7%, respectively, in the British and Italian patients (see Materials and methods, Supplementary Material, Table S10). Our results confirm data previously reported (17–20). We performed a meta-analysis of this sample subset (2287 sporadic cases and 4162 controls) including and excluding cases (n = 144) with the expanded hexanucleotide repeat. Significant association of SNP rs3849943 decreased from P = 7.72 × 10−5 to P = 0.052 when the carriers were excluded from the logistic regression analysis, confirming that the association of 9p21 locus was largely dependent on carriers of the C9orf72 expansion (Supplementary Material, Fig. S4).

**DISCUSSION**

In this study, we present the largest GWAS meta-analysis to date in ALS. Overall, we analysed 13 225 individuals with sufficient
power to capture allelic association with small effects and low MAF. The novel associated SNP rs34517613 at 17q11.2 was confirmed when genotype data from the independent replication cohort were combined in meta-analysis with full GWAS data. We have also replicated the 9p21 locus at genome-wide significance; this locus was also identified by linkage studies of familial ALS patients and by analogy it may be that there is a corresponding familial locus for the 17q11.2 locus identified here, though we are not aware of any reports.

The lead SNP at this locus was in LD with three SNPs in the SARM1 gene. Interestingly, SARM1 orthologues, dSarm in Drosophila melanogaster and Sarml in mice, have been recently found to play a direct role in an axonal self-destruction pathway, a mechanism known as Wallerian degeneration, which shares morphological similarities with the axon dying-back degeneration observed in ALS and in other neurodegenerative pathologies (30,31). As axonal degeneration is an early feature of ALS progression, a mouse model for familial ALS, SOD1G93A, was crossed with the WldS mouse, a spontaneous mutant with phenotype of prolonged survival of injured axons because of a chromosomal rearrangement that disrupts two genes, Ube4b and Nmnat1. In the SOD1G93A/WldS, the progression of axonopathies was attenuated, suggesting that the Wallerian pathway could be involved in the axon loss observed in ALS (32,33).

We estimated the heritability of ALS owing to common variation as ~12% based on a prevalence of 5 per 100 000 (1,2), although the true prevalence of ALS may be higher as consequence of increasing age in the human population as discussed earlier. We have also only considered autosomal SNPs and additional variance may be encoded on the X chromosome. Twin studies estimate heritability owing to all genetic variation including rare monogenic forms of the disease, whereas we have estimated heritability as a consequence of common polymorphisms either directly typed or imputed. The difference between heritability estimated from twin studies and from analysis of common SNP polygenic variation suggests a substantial role for variation not captured by genome-wide association studies.

Our estimate for the heritability of ALS explained by common variation is lower than that for other late-onset neurodegenerative diseases. Using similar approaches, heritability for late-onset Parkinson’s disease was 0.31 (34), 0.24 for late-onset Alzheimer’s disease and 0.30 for multiple sclerosis (35). These differences in heritability owing to common variants are reflected in the difficulty in identifying genome-wide significant loci for ALS compared with these diseases. For example, a two-stage GWA study of Parkinson’s disease with a stage 1 sample size of ~6000 individuals identified two strong association signals (36), and a study of Alzheimer’s disease with a similarly sized stage 1 sample identified two loci additional to the well-replicated APOE locus (37). In comparison with these studies, our combined sample size is >13 000 individuals, suggesting that the genetic architecture of ALS may be different from other more common neurodegenerative diseases. A recent study of ALS in a Han Chinese population identified two loci of genome-wide significance in a combined sample of ~5 000 (38); strikingly these loci were not replicated in populations of European ancestry, nor in the combined meta-analysis reported here. This suggests there may be considerable heterogeneity in ALS risk loci across different ethnicities; in support of this, the expanded repeat in the C9orf72 gene has been reported to be infrequent in Asian populations (39,40).

In conclusion, we have identified a novel locus for sporadic ALS risk at 17q11.2, but not replicated the suggestive evidence for a second locus at 18q11.2 and confirmed the association at 9p21 with the expanded hexanucleotide repeat in the C9orf72 gene. Furthermore, we provide evidence from our heritability estimates that further common variation affecting ALS risk remains to be detected by current GWAS platforms and with larger cohorts but that denser genome-wide assays and next generation sequencing technologies are required to detect rarer variation.

MATERIALS AND METHODS

Participating individuals

The discovery sample consisted of a novel Italian cohort collected by the Italian Consortium for the Genetics of ALS (SLAGEN) and seven independent published studies collected by ALSGEN, The International Consortium on Amyotrophic Lateral Sclerosis Genetics (5–11,15,16). All patients fulfilled the El Escorial revised criteria for ALS (41). Written informed consent according to the Declaration of Helsinki was obtained from all patients and healthy subjects participating in this study. Local ethical committees for each participating institution approved these studies. All samples across the eight datasets were of European ancestry, and a full description of demographic details is reported in the Supplementary Material, Table S1. All patients with family history of ALS or carrying Mendelian risk genes were excluded from analyses. Screening for the expanded hexanucleotide repeat in the C9orf72 gene were performed subsequently; information of mutation carriers was available for a subset of cases included in the Italian, Dutch and British cohorts (see Supplementary Material for full description).

A replication sample of 4630 individuals (2074 cases and 2556 controls) was collected across Italy, The Netherlands and Germany using the same diagnostic criteria (see Supplementary Material, Table S8).

Genotyping procedures

Samples from the individual cohorts were genotyped on different Illumina beadchips (Illumina, CA, USA) as shown in Table 1. Additional SNPs were genotyped in the replication phase by PCR-based KASP technique by the KBioscience (UK) facility for the Italian samples (622 cases and 971 controls) or by TaqMan 2013 PCR method for the Dutch (877 cases and 1226 controls) and German samples (575 cases and 359 controls) (Supplementary Material). Cross-validation between methodologies showed 100% concordance.

Statistical analysis

Quality control and meta-analysis of imputed data

Before imputation analysis QC of samples and markers was performed separately in the eight studies using the PLINK software package (42). A detailed description of QC procedure is reported in the Supplementary Material and Table S2. Genotype data
from British cases and controls were merged and analysed as single cohort as fully described in the Supplementary Material and Table S2.

Ancestry differences between individuals within each cohort were detected by PCA. Principal component axes were generated by genotypes of a genome-wide subset of LD-independent SNPs using EIGENSTRAT software (24,25), and outliers identified by the first 10 principal components (PCs) were removed. The number of significant principal components was estimated by Tracy–Widom distribution using the program TW statistic (25) and included as covariates in the logistic regression analyses of each study for population stratification control. Scatter plots of PC1 and PC2 showed no evidence of substructure between cases and controls within the single cohorts (Supplementary Material, Fig. S1). As an example, Supplementary Material, Figure S5 shows the population structure of the novel Italian cohort by the projection of the first two PC axes.

After QC, original genotype data of each study were tested for genomic inflation and lambda estimates resulted to be minimal ($\lambda_{(gc)} < 1.02$). Individual datasets were imputed genome wide separately using the IMPUTE.v2 program (43,44), which employs combined reference panels of known phased haplotypes provided by HapMap 3 (Feb 2009), 1000 Genomes Project (Mar 2010) (NCBI build 36 coordinates) and the study’s sample genotypes (Supplementary Material). In each cohort, imputed genotypes were tested for association with ALS status by logistic regression analysis (SNPTESTv2.4.0) including the specific PCs as confounder covariates to control population stratification. For each test, statistic spurious associations and genomic inflation were controlled by plotting the observed quantities versus the expected ones in Q–Q plots and by factor $\lambda_{(gc)}$ estimate (Supplementary Material, Fig. S7 and Table S3) (R package software).

SNP statistic data were filtered for uncertainty of inferred genotypes according to posterior probability (APP) of $>0.9$ and statistical information of allele frequency (info) of $>0.4$ (26). The average number of SNPs filtered out in each dataset was 1.47 million (Supplementary Material, Table S2). Finally, filtered SNP tables were combined in a meta-analysis and analysed with the program METAL (27), applying the standard error scheme option that weights effect size estimates, or the program METAL (27), applying the standard error scheme.

The variance estimate of each cohort was combined by meta-analysis (METAL) applying standard error scheme of the random effects meta-analysis weighted by sample size using the program $rmeta$ in R toolset.

Power calculation

The final meta-analysis cohort had 99.99% power to detect allelic association with an odds ratio (OR) of 1.2 and MAF of 0.25 at genome-wide significance ($P = 5 \times 10^{-8}$). For low-frequency variants with MAF ranging from 0.01 to 0.04 and of small effect (OR $< 1.5$), there was low power to capture variants with MAF 0.01, whereas for SNPs with MAF 0.02 power ranged from 60 to 98% for alleles with OR of 1.3 and 1.4, respectively (Supplementary Material, Fig. S8).

Screening of expanded repeats in C9orf72 gene

Hexanucleotide repeat expansion data for the C9orf72 gene were available for the Italian, Dutch and British cases (Supplementary Material, Table S10). As previously reported, carriers were defined as individuals with a range of the GGGGCC repeats of $>23$ units (17–20). We performed conditional logistic regression analysis on a subset of 2287 screened cases and 4162 controls including and excluding pathologic expansions carriers ($n = 144$).

Estimation of the genetic variance tagged by all SNPs

We used GCTA software (29) to estimate the proportion of ALS phenotypic variance explained by autosomal SNPs distributed genome wide. Genetic relationships between pairs of individuals were calculated separately in the eight cohorts where all samples were previously tested for cryptic relatedness and excluded if the proportion of IBD (identical-by-descent) estimate was $>0.05$ (Supplementary Material, Table S2). We estimated the genetic relationship matrix (GRM) including imputed genotypes filtered by MAF $>0.01$, posterior probability $>0.9$ and information measure $>0.8$. Filtered genotype data were submitted to GRM analysis, and GRMs output data were used in the REML analysis. In each dataset, REML analysis was carried out by fitting the specific principal components (Supplementary Material, Table S3) as covariates to control for possible population stratification. The variance estimate was transformed from the observed scale ($V(1)/Vp$) to a scale of liability ($V(1)/V_{p,L}$) by specifying a disease prevalence for ALS of 5 per 100 000 persons (1,2). The variance estimate of each cohort was combined by random effects meta-analysis weighted by sample size using the library $rmeta$ in R toolset.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

DATA AVAILABILITY

Data were accessed by formal request to the consortia members. The complete summary meta-analysis data (SNP, genomic position, odds ratio and $P$-value) are freely available at the ALSOD website (http://alsod.iop.kcl.ac.uk/) in a searchable format (47).
ACKNOWLEDGEMENTS

Genetic and statistical analysis was performed using a Linux cluster supported by the National Institute for Health Research Biomedical Research Centre for Mental Health at the South London and Maudsley National Health Service Foundation Trust and Institute of Psychiatry King’s College London. G.So. acknowledges EuroBiobank for ALS DNA samples; K.M. acknowledges Motor Neurone Disease Association UK, Queen Elizabeth Hospital Birmingham Charity. M.S. acknowledges support through EU via the Euro-MOTOR consortium and German BMBF Motoneuron disease network.

Conflict of Interest statement: None declared.

FUNDING

The SLAGEN Consortium was supported by funds from the Ministero della Salute Italiano (GWA-SLA, Ricerca Finalizzata 2007 grant number 31) and Fondazione Istituto Auxologico Italiano; I.F.C., C.G., B.C., J.F.P. and V.S. were supported by the Istituto Superiore di Sanità (grant number 526D34, 2006); I.F. received salary support from ‘Amici del Centro Dino Ferrari’; V.S. acknowledges support from the ‘Amici del Centro Dino Ferrari’. A.R., C.G., B.C., N.T. and V.S. were supported by the Ministero della Salute Italiano (ALS-FTD, RF-2009-1473856); G.Si. was supported by Monte dei Paschi di Siena Foundation (grant 39167) and the Ministero della Salute Italiano (grant number RF-INP-2007-627227); A.C. was supported in part by Federazione Italiana Giuoco Calcio and Fondazione Vialli e Mauro per la Sclerosi Laterale Amiotrofica onlus, Ministero della Salute Italiano (grant number RF-PIE-2007-635695); AF was supported by Progetto Regione Emilia Romagna-Università ‘RARER’. The Italian replica control samples were obtained from the ‘Parkinson Institute Biobank’ and the University Attraction Poles (IUAP) program P7/16 of the University of Leuven (grant number GOA/11/014), the International Trust and Institute of Psychiatry, King’s College London.

RADIANT depression and bipolar case control studies (DeCC, BACC) were genotyped under Medical Research Council (grant number G0701420). A.A.-C. thanks the Motor Neurone Disease Association of Great Britain and Northern Ireland, the ALS Association, the ALS Therapy Alliance and the Medical Research Council (grant number G0600974) for support. R.H.B. acknowledges support from the National Institute of Neurological Disorders and Stroke (grant number 5RO1NS050557-05) and the National Institute of Neurological Disorders and Stroke American Recovery and Reinvestment Act Award (grant number RC2-NS070-342), the Angel Fund, the ALS Association, P2ALS, Project ALS, the Pierre L. de Bourgknecht ALS Research Foundation and the ALS Therapy Alliance. J.E.L. acknowledges grant support from National Institute of Health/ National Institute of Neurological Disorders and Stroke (grant number 1R01NS073873). The Dutch, Belgian and Swedish GWAS data generation was supported by the Princes Beatrice Fonds, VSB Fonds, H. Kersten and M. Kersten (Kersten Foundation), The Netherlands ALS Foundation, and J.R. van Dijk and the Aedsessum Foundation.

REFERENCES

London, UK; and Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy), Gellera Cinzia (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Ratti Antonia (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano; and Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, University degli Studi di Milano, Milano, Italy), Silani Vincenzo (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano; and Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, Università degli Studi di Milano, Milano, Italy), Sorari Gianni (Department of Neurosciences, University of Padova, Padova, Italy).

Slagen collaborators

Bongioanni Paolo (Neurorehabilitation Unit, Azienda Ospedaliero-Universitaria Pisana, Pisa, Italy), Brescia Morra Vincenzo (Department of Neurological Sciences, University Federico II, Napoli, Italy), Calini Daniela (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy), Castellotti Barbara (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Ceroni Mauro (Laboratory of Experimental Neurobiology, IRCCS ‘C. Mondino’ National Institute of Neurology Foundation, Pavia, Italy), Conforti Francesca Luisa (Institute of Neurological Sciences, National Research Council, Mangone, Italy), Corbo Massimo (NeuroMuscular Omnicentre-NEMO, Fondazione Serena Onlus, Niguarda Ca’ Granda Hospital, Milano, Italy), Corrado Lucia (Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases, ‘A. Avogadro’ University, Novara, Italy), Corti Stefania (Neurologic Unit, IRCCS Foundation Ca’ Granda Ospedale Maggiore Policlinico; and Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, Università degli Studi di Milano, Milano, Italy), Del Bo Roberto (Neurologic Unit, IRCCS Foundation Ca’ Granda Ospedale Maggiore Policlinico; and Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, Università degli Studi di Milano, Milano, Italy), Ferlini Alessandra (Department of Medical Sciences, University of Ferrara, Ferrara, Italy), Filla Alessandro (Department of Neurological Sciences, University Federico II, Napoli, Italy), Filosto Massimo (Section for Neuromuscular Diseases and Neuropathies, Clinical Neurology, University Hospital ‘Spedali Civili’, Brescia, Italy), Gagliardi Stella (Laboratory of Experimental Neurobiology, IRCCS ‘C. Mondino’ National Institute of Neurology Foundation, Pavia, Italy), Inghilleri Maurizio (Department of Neurology and Psychiatry, National University of Rome ‘Sapienza’, Roma, Italy), Marsili Angela (Department of Neurological Sciences, University Federico II, Napoli, Italy), Mazzini Letizia (Department of Neurological Sciences, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Pensato Viviana (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Purozzo Giorgia (Department of Neurological Sciences, University Federico II, Napoli, Italy), Querin Giorgio (Department of Neurosciences, University of Padova, Padova, Italy), Saccà Francesco (Department of Neurological Sciences, University Federico II, Napoli, Italy), Siciliano Gabriele (Division of Neurology, Department of Clinical and Experimental Medicine, University of Pisa, Nuovo Ospedale S.Chiara-Cisanello, Pisa, Italy), Sorbi Sandro (Department of Neurological and Psychiatric Sciences, University of Firenze, Firenze, Italy), Taroni Franco (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Ticozzi Nicola (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano; and Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, Università degli Studi di Milano, Milano, Italy), Tiloca Cinzia (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy).

Collaborators

La Bella Vincenzo (ALS Clinical Research Center, Department of Experimental Biomedicine and Clinical Neurosciences, Bio.Ne.C, University of Palermo, Palermo, Italy), Logroscino Giancarlo (Department of Neurosciences and Sense Organs, University of Bari, Policlinico, Bari, Italy), Monsuor María Rosaria (Second Division of Neurology, Second University of Naples, Naples, Italy), Quatrone Aldo (Institute of Neurology, University Magna Graecia, Catanzaro, Italy), Simone Isabella Laura (Department of Neurosciences and Sense Organs, University of Bari, Policlinico, Bari, Italy).

Alsgen consortium collaborators

Ahmeti Kreshnik B. (Northwestern University School of Medicine, Davee Department of Neurology, Division of Neuromuscular Medicine, Chicago, IL, USA), Ajroud-Driss Senda (Northwestern University School of Medicine, Davee Department of Neurology, Division of Neuromuscular Medicine, Chicago, IL, USA), Al-Chalabi Ammar (MRC Centre for Neurodegeneration Research, Institute of Psychiatry, King’s College London, London, UK), Andersen Peter M. (Institute of Pharmacology and Clinical Neuroscience, Section for Neurology, Umeå University, Umeå, Sweden), Armstrong Jennifer (Northwestern University School of Medicine, Davee Department of Neurology, Division of Neuromuscular Medicine, Chicago, IL, USA), Birve Anne (Institute of Pharmacology and Clinical Neuroscience, Section for Neurology, Umeå University, Umeå, Sweden), Blauw Hylke M. (Rudolf Magnus Institute of Neuroscience, Department of Neurology, University Medical Centre Utrecht, Utrecht, The Netherlands), Brown Robert H. (Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA.), Brujin Lucie (The ALS Association), Chen Wenjie (Northwestern University School of Medicine, Davee Department of Neurology, Division of Neuromuscular Medicine, Chicago, IL, USA), Chiò Adriano (ALS Center, Department of Neurosciences, University of Turin, Torino, Italy), Comeau Mary C. (Wake Forest University Health Sciences, Medical Center Blvdr, Winston-Salem, NC, USA), Cronin Simon (Department of Neurology, Beaumont Hospital, Dublin, Ireland), Diekstra Frank P. (Rudolf Magnus Institute of Neuroscience, Department of Neurology, University Medical Centre Utrecht, Utrecht, The Netherlands), Gellera Cinzia (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Logroscino Giancarlo (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy), Pensato Viviana (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Purozzo Giorgia (Department of Neurological Sciences, University Federico II, Napoli, Italy), Querin Giorgio (Department of Neurosciences, University of Padova, Padova, Italy), Saccà Francesco (Department of Neurological Sciences, University Federico II, Napoli, Italy), Siciliano Gabriele (Division of Neurology, Department of Clinical and Experimental Medicine, University of Pisa, Nuovo Ospedale S.Chiara-Cisanello, Pisa, Italy), Sorbi Sandro (Department of Neurological and Psychiatric Sciences, University of Firenze, Firenze, Italy), Taroni Franco (Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico ‘Carlo Besta’, Milano, Italy), Ticozzi Nicola (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano; and Department of Pathophysiology and Transplantation, ‘Dino Ferrari’ Center, Università degli Studi di Milano, Milano, Italy), Tiloca Cinzia (Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy).