Human Molecular Genetics, 2008, Vol. 17, No. 5 ***768–774***
doi:10.1093/hmg/ddm361
Advance Access published on December 15, 2007

A genome-wide association study of sporadic ALS in a homogenous Irish population

Simon Cronin[1,2*], Stephen Berger[3], Jinhui Ding[4], Jennifer C Schymick[3,7], Nicole Washecka[3], Dena G. Hernandez[3], Matthew J. Greenway[2], Daniel G. Bradley[9], Bryan J. Traynor[3,5,6,8] and Orla Hardiman[2,10]

[1]Department of Clinical Neurological Sciences, Royal College of Surgeons in Ireland, Dublin 2, Ireland, [2]Department of Neurology, Beaumont Hospital, Dublin 9, Ireland, [3]Laboratory of Neurogenetics, National Institute on Aging, [4]Computational Biology Core, Laboratory of Neurogenetics, National Molecular Genetics Unit, [5]Neurogenetics Branch, National Institute of Neurological Disorders and Stroke and [6]Molecular Genetics Unit, National Institutes of Health, Bethesda, MD, USA, [7]Department of Physiology, Anatomy and Genetics, University of Oxford, Henry Wellcome Building of Gene Function, Oxford, UK, [8]Neurology Department, Johns Hopkins University, Baltimore, MD, USA, [9]Smurfit Institute of Genetics and [10]Trinity College Institute of Neuroscience, Trinity College, Dublin 2, Ireland

Received November 10, 2007; Revised and Accepted December 4, 2007

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by progressive limb or bulbar weakness. Efforts to elucidate the disease-associated loci have to date produced conflicting results. One strategy to improve power in genome-wide studies is to genotype a genetically homogenous population. Such a population exhibits extended linkage disequilibrium (LD) and lower allelic heterogeneity to facilitate disease gene mapping. We sought to identify associated variants for ALS in the Irish, a stable population of relatively homogenous genetic background, and to replicate these findings in larger genetically out-bred populations. We conducted a genome-wide association study in 432 Irish individuals using Illumina HumanHap 550K single nucleotide polymorphism chips. We demonstrated extended LD and increased homogeneity in the Irish sample when compared to an out-bred population of mixed European ancestry. The Irish scan identified 35 loci associated with *P*-values below 0.0001. For replication, we identified seven chromosomal regions commonly associated in a joint analysis of genome-wide data on 958 ALS cases and 932 controls from Ireland and the previously published datasets from the US and The Netherlands. When pooled, the strongest association was a variant in the gene encoding *DPP6*, a component of type A neuronal transmembrane potassium channels. Further confirmation of the candidate loci is warranted in additional genome-wide datasets. We have made our individual genotyping data publicly available, contributing to a powerful world-wide resource to refine our understanding of the genetics of sporadic ALS.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterized by adult-onset loss of motor neurons. The peak of onset is in the fifth to seventh decades, and average survival from symptom onset is 3–5 years. Despite extensive efforts to identify the aetiology and potential treatments for ALS, the only disease modifying agent for which there is compelling evidence is riluzole, which increases survival by ~3–6 months (1).

Population-based epidemiological studies indicate that 2–5% of cases are associated with familial, commonly autosomal dominant, inheritance (2). For ~20% of familial ALS, mutations have been identified in genes, including Cu/Zn

*To whom correspondence should be addressed at: The Irish ALS Research Group, Department of Neurology, Beaumont Hospital, Dublin 9, Ireland. Tel: +353 18092174; Fax: +353 18092302; Email: scronin@rcsi.ie

superoxide-dismutase, dynactin 1, alsin, vesicle-associated protein B, senataxin and angiogenin (3–8). The cause of sporadic ALS is less well understood, although it is widely believed that genetic factors play a central role. Despite this, attempts to identify genetic variants associated with sporadic ALS using candidate gene approaches have produced heterogeneous and often disappointing results. (9).

Recent advances, such as the International HapMap Project (10) and the development of robust, high-throughput genotyping platforms provide the opportunity to rapidly screen common genetic variation across the human genome for association with disease. In contrast to candidate gene studies, genome wide association studies are unbiased by *a priori* hypotheses about disease biology. Indeed, genome wide association studies have identified genetic susceptibility factors for common diseases, including heart disease, diabetes and cancer (11–13). However, these studies utilized several thousand cases and controls to achieve adequate statistical power, whereas the lower incidence of ALS among the general population limits sample availability. Accordingly, initial reports of genome-wide association (GWA) study data for sporadic ALS have produced heterogeneous associations to differing genetic loci, owing to relatively smaller sample numbers assayed in out-bred populations (14–16). One strategy to increase power in association studies is to genotype a genetically homogenous population. Such a population exhibits extended linkage disequilibrium (LD) and lower allelic heterogeneity, which combine to increase tagging efficiency, lower false positive association rate and harmonize the background genetic structure of the study participants (17–18).

Geography and history have contributed to the relatively simple structure of the Irish gene pool (19). Ireland's position on the western edge of Europe means that its population has been relatively undisturbed by major demographic movements across the continent. Ireland was founded by Mesolithic settlers in ~7000BC, possibly followed by further settlers in Neolithic and later periods. There is evidence of restricted movement of peoples throughout much of the island's history, such as laws forbidding travel outside of small fiefdoms or *tuath*, and these patterns continued in rural areas until recent times (20). Several lines of evidence have identified a shared genetic affinity along the Atlantic façade of Europe, which perhaps reflects the genetic signature of ancient settlers (21). The strength of this signal, for example the high prevalence of the R1b3 Y chromosome haplotype in the present day Irish population, provides evidence of the limited admixture that has occurred within the island. (19).

We hypothesized that the relatively homogenous genetic structure of the Irish population simplifies the identification of genetic loci that alter risk of developing ALS. Here, we report a whole-genome association study involving 221 Irish patients with ALS and 211 neurologically normal controls in which 540 466 single nucleotide polymorphisms (SNPs) were genotyped using the Illumina HumanHap550 SNP chips. We compare LD between the Irish and a US population, ostensibly an out-bred population of mixed European ancestry. To replicate our findings, we compare all associated SNPs with previously published genome-wide data from the United States and The Netherlands (15–16).

RESULTS

We genotyped 439 unique Irish individuals consisting of 222 patients with sporadic ALS and 217 control subjects using the Infinium HumanHap550 SNP chip (Illumina Inc., San Diego). The mean call rate across all samples was 99.8% and the lowest call rate was 97.7%. Of the 561 466 SNPs assayed, 20 913 were excluded due to a call rate below 98% or cluster separation below 0.3, which is a marker of genotyping accuracy generated by the BeadStudio v3.1 software (Illumina Inc.). Among the remaining 527 279 autosomal SNPs, the call rate was above 99% for 516 911 SNPs and above 98% for all. 29 362 SNPs were then excluded as they had a minor allele frequency (MAF) below 0.01 ($n = 23\ 060$) and/or a Hardy–Weinberg equilibrium (HWE) *P*-value below 0.01 in controls ($n = 6340$). The final association analysis was conducted using the remaining 497 917 autosomal SNPs.

One patient and six controls were excluded from further analysis as follows: in one patient (sample ID = Irish_348) and one control (ID = Irish_30m), X chromosome heterozygosity did not match recorded gender; four controls (sample IDs Irish_59c, Irish_51f, Irish_27c and Irish_140 m) were identified as being cryptically related based on high identity-by-state similarity scores generated within the PLINK 0.99s software package (22) (threshold set as score greater than 75%); finally, a control individual (sample ID = Irish_77c) was identified as having probability of Asian ancestry using the STRUCTURE 2.1 software (Supplementary Material, Fig. S1) (23). STRUCTURE and Eigenstrat 2.0 analyses (24) of the remaining 432 individuals indicated no discernible population substructure between cases and controls (Supplementary Material, Fig. S2). After these exclusions, the final cohort consisted of 221 cases and 211 controls, all of which were included in the association analysis (Table 1).

We next compared allelic heterogeneity and LD between the Irish population and the previously published US cohort, representing an out-bred population of European ancestry (15). To compare allelic heterogeneity, we calculated minor allele frequencies for all SNPs genotyped in both populations. We observed 59.4% more monomorphic SNPs in the Irish than the out-bred US sample, and minor allele frequencies were lower for rare polymorphisms. To compare the extent of LD between the Irish and out-bred samples, we calculated the inter-marker r^2 statistic between the same pairs of SNPs in each population, and then grouped comparisons into distance bins according to increasing inter-marker separation. We used two measures to compare LD data between the populations. First, we compared mean pair-wise r^2 for the two populations in each 100 kb distance bin. The mean r^2 was higher in the Irish population at all distances out to 1 Mb. Between 100 and 600 kb mean r^2 was higher in the Irish by an average of 5.5%. Thereafter, the difference in mean r^2 declined with increasing inter-marker separation (Fig. 1A). Second, to assess the potential of this pattern for capturing non-genotyped disease-associated variants, we measured the number of SNP pairs with highly correlated minor alleles ($r^2 > 0.8$) within all distance bins of 40 kb (25). We observed between 6 and 12% more tagging pairs in the Irish population compared to the US population at distances out to 280 kb. After this point the difference between the two populations

Table 1. Characteristics the Irish, US and The Netherlands study populations

	Total	Male (%)	Female (%)	Age at onset (years, mean)	Spinal onset (%)	Bulbar onset (%)
Irish 550 k GWAS						
Patients with sporadic ALS	221	54	46	61	72	28
Controls	211	53	47	58		
United States 550K GWAS						
Patients with sporadic ALS	276	63	37	55	76	24
Controls	271	48	52	68		
The Netherlands 300K GWAS						
Patients with sporadic ALS	461	59	41	59	69	31
Controls	450	59	41	60		

GWAS: genome-wide association study.

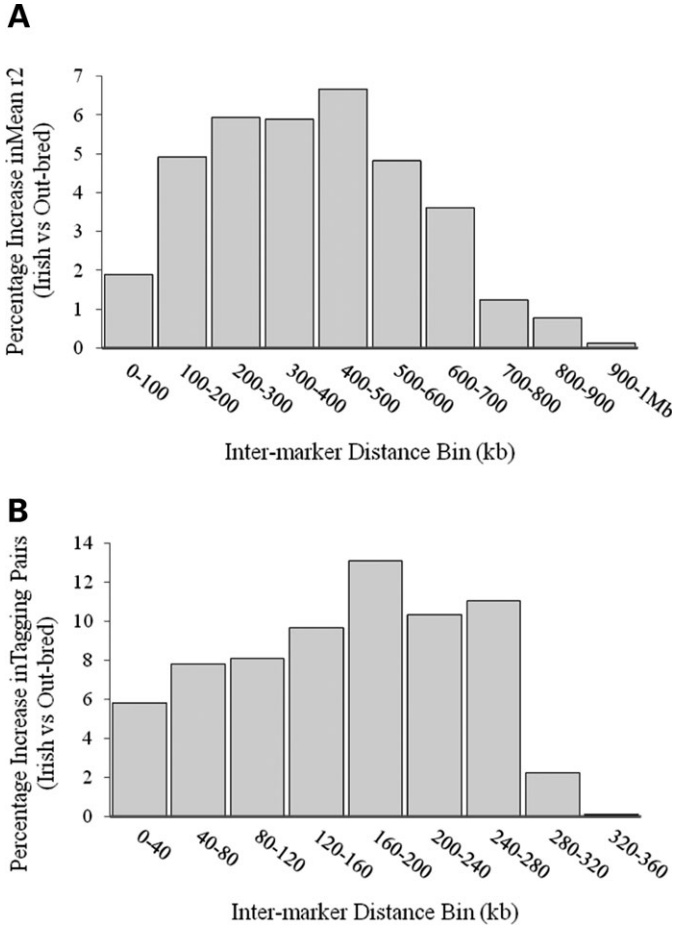

Figure 1. Increase in linkage disequilibrium (LD) in the Irish population compared to an outbred population of European ancestry. LD was measured by pairwise comparison of r^2 between all markers that fell into the same inter-marker distance bin. (**A**) Percentage increase in mean r^2 and (**B**) percentage increase in highly correlated inter-marker comparisons ($r^2 > 0.8$).

declined as the number of highly-correlated pairs became smaller (Fig. 1B). No difference in LD pattern was observed between patients with ALS and controls.

Allelic association tests were performed between the 221 Irish ALS cases and 211 Irish controls using 497 917 autosomal SNPs that had a MAF of 0.01 or more, a call rate of 98% or more and probability of deviation from HWE of 0.01 or more. Table 2 shows results for 35 SNPs that were associated with ALS at a *P*-value less than 0.0001 in the Irish. Among these, marker redundancy accounted for three linked SNPs on chromosome 13 (rs7316983, rs3813131, rs3813133, inter-marker $r^2 > 0.98$), and two SNPs on chromosomes 5 (rs252095 with rs252139, $r^2 = 0.99$).

To validate our allelic association findings, we compared the Irish cohort ($n = 221$ cases and 211 controls) with recently published genome wide association studies performed on a cohort of 276 US ALS cases and 271 US controls (15) and on a cohort of 461 Dutch ALS patients and 450 controls (16). The Irish and US cohorts were both genotyped on Infinium HumanHap550 SNP chips (genotypes>550 000 SNPs), whereas the Dutch cohort was genotyped on Infinium HumanHap317 SNP chips, which genotype 317 511 SNPs. There were 287 522 autosomal SNPs in common between these two platforms that were successfully genotyped in all three cohorts. Of these, there were nine SNPs common to all three datasets, associated for the same allele at a *P*-value less than 0.05 (Table 3). The strongest association was for rs10260404, an intronic variant lying within the *DPP6* gene on chromosome 7 (allelic *P*-value for all three cohorts$=2.53 \times 10^{-6}$, odds ratio = 1.37, 95% CI: 1.2–1.56).

DISCUSSION

We have conducted a GWA study in a genetically homogenous Irish population of 221 ALS cases and 211 controls. LD mapping confirmed increased homogeneity within our case-control cohort, as compared to an out-bred population. Most importantly, comparison of our data on a genome-wide scale with replication series from the United States and The Netherlands has identified seven genomic regions showing significant association in all three populations at a *P*-value less than 0.05.

Our central hypothesis was that the isolated and genetically homogeneous nature of the Irish population facilitates disease gene mapping because of both the increased extent of LD and the reduced environmental and genetic heterogeneity observed in such groups (18,26). Several examples exist of true population isolates, cut off for centuries with high endogamy in rural villages and regions such as Talhana (Sardinia), Kuusamo (Finland) and Kosrae (Micronesia) (25,27,28). While the potential of these isolates for mapping common diseases has been clearly demonstrated (25), their small population size precludes study of rare conditions, such as ALS, which has an incidence of ~2.1 per 100,000 person-years in the general population (29). In this situation, larger but homogeneous populations, in which genetic heterogeneity has been reduced as a result of founder effects and cultural and linguistic isolation, may prove a useful alternative (18). Indeed, analysis of the genotype data generated as part of this project supports the notion that the Irish population is genetically homogeneous: we found higher LD in our unselected Irish

Table 2. Single nucleotide polymorphisms (SNPs) associated with sporadic ALS at uncorrected allelic *P*-value < 0.0001 in the Irish study population

Chr	SNP	Location (build 36.2)	Gene	Putative function	HWE *P*-value	MAF SALS	Control	Allelic association *P*-value
13	rs7316983	114092980	LOC2834	Uncharacterized	0.72	0.10	0.02	1.05×10^{-5}
13	rs3813131	114108121	LOC2834	Uncharacterized	0.72	0.10	0.02	1.05×10^{-5}
13	rs1325803	42103269	Intergenic		0.53	0.38	0.24	1.57×10^{-5}
13	rs3813133	114108295	LOC2834	Uncharacterized	0.72	0.10	0.02	1.67×10^{-5}
1	rs1915250	235509815	Intergenic		0.36	0.12	0.05	3.05×10^{-5}
17	rs1558878	63876399	ARSG	Arylsulphatase G; missense Arg>Trp mutation; cell signalling; hydrolyzation of sulphate esters and degradation of macromolecules	0.48	0.42	0.56	3.50×10^{-5}
1	rs11208807	66346942	PDE4B	Phosphodiesterase 4B; intra-cellular second messenger to hormones and neurotransmitters	0.17	0.25	0.38	3.63×10^{-5}
8	rs10106208	122814511	Intergenic		0.38	0.09	0.18	3.88×10^{-5}
9	rs17724552	107079909	Intergenic		0.75	0.08	0.18	3.92×10^{-5}
14	rs712436	47521184	Intergenic		0.37	0.36	0.50	4.21×10^{-5}
18	rs7245160	70417826	LOC400657	Hypothetical protein	0.87	0.11	0.21	4.22×10^{-5}
8	rs6473902	55080776	TCEA1	Transcription elongation factor A1; cofactor with RNA polymerase	0.97	0.11	0.21	4.29×10^{-5}
5	rs252095	141342346	RNF14	Ring finger protein 14; androgen receptor co-activator, ubiquitination	0.74	0.11	0.22	4.58×10^{-5}
5	rs409037	40373686	Intergenic		0.74	0.11	0.22	4.58×10^{-5}
16	rs1551960	7082457	A2BP1	Ataxin 2-binding protein 1; *trans*-Golgi network protein; implicated in spinocerebellar ataxia type 2 (SCA2)	0.62	0.05	0.14	4.67×10^{-5}
6	rs9328053	1389294	FOXF2	Forkhead box F2; transcription factor	0.22	0.28	0.41	4.70×10^{-5}
2	rs4672448	61839751	Intergenic		0.5	0.17	0.08	4.82×10^{-5}
4	rs1013284	14065624	Intergenic		0.76	0.25	0.38	4.93×10^{-5}
8	rs7817815	5532399	Intergenic		0.67	0.21	0.11	4.96×10^{-5}
13	rs2408213	50350752	Intergenic		0.92	0.35	0.22	5.24×10^{-5}
1	rs17105335	49148127	ABGL4	Hypothetical protein	0.83	0.02	0.08	5.42×10^{-5}
11	rs873108	76133266	Intergenic		0.94	0.11	0.21	5.54×10^{-5}
13	rs9512144	25657079	RNF6	Ring finger protein 6; regulation of transcription	0.06	0.31	0.45	6.15×10^{-5}
4	rs4640677	103015253	Intergenic		0.48	0.10	0.19	6.62×10^{-5}
14	rs2770409	47468271	Intergenic			0.33	0.46	6.92×10^{-5}
5	rs252139	141325402	RNF14	Ring finger protein 14; androgen receptor co-activator, ubiquitination	0.74	0.11	0.22	7.04×10^{-5}
13	rs4772972	107908415	Intergenic		0.16	0.20	0.32	7.05×10^{-5}
18	rs4798376	5594240	Intergenic		0.07	0.32	0.20	7.57×10^{-5}
4	rs5019445	182501271	Intergenic		0.76	0.34	0.22	7.58×10^{-5}
12	rs950795	125816883	Intergenic		0.78	0.21	0.33	8.00×10^{-5}
10	rs17527491	23067587	Intergenic		0.74	0.34	0.22	8.44×10^{-5}
13	rs17605645	68778147	Intergenic		0.13	0.19	0.09	9.21×10^{-5}
2	rs2374482	43217577	Intergenic		0.39	0.44	0.31	9.73×10^{-5}
4	rs12648641	103016928	Intergenic		0.62	0.06	0.14	9.89×10^{-5}

Chr: chromosome; SNP: single nucleotide polymorphism; HWE: Hardy-Weinberg equilibrium; MAF: minor allele frequency; SALS: sporadic ALS.

Table 3. Minor allele frequencies and pooled *P*-values for the nine SNPs associated with SALS independently in the Irish, US and The Netherlands study populations at a *P*-value less than 0.05

Chr	SNP	Location (build 36)	Gene	Ireland MAF		United States MAF		Netherlands MAF		Pooled *P*-value	OR (95% CI)
				SALS	Control	SALS	Control	SALS	Control		
7	rs10260404	153648446	DPP6	0.42	0.34	0.43	0.34	0.44	0.37	2.53×10^{-6}	1.37 (1.2–1.56)
16	rs17823157	12584030	Intergenic	0.26	0.2	0.31	0.23	0.28	0.24	1.42×10^{-5}	1.38 (1.2–1.6)
18	rs2159942	14052349	Intergenic	0.26	0.19	0.25	0.19	0.29	0.24	3.96×10^{-5}	1.37 (1.18–1.59)
10	rs7899260	12091580	UPF2	0.09	0.14	0.07	0.11	0.07	0.11	4.04×10^{-5}	1.58 (1.27–1.97)
4	rs6851312	175542883	FBXO8	0.34	0.42	0.35	0.42	0.32	0.38	7.34×10^{-5}	1.31 (1.14–1.49)
4	rs6824643	175542813	FBXO8	0.34	0.42	0.35	0.42	0.32	0.38	7.57×10^{-5}	1.31 (1.14–1.49)
4	rs2251316	175833019	Intergenic (near FBXO8)	0.26	0.33	0.25	0.31	0.32	0.37	1.02×10^{-4}	1.31 (1.14–1.51)
3	rs6779441	119498923	Intergenic	0.08	0.05	0.07	0.04	0.08	0.05	1.35×10^{-4}	1.69 (1.29–2.22)
20	rs458032	15167500	LOC140733	0.34	0.27	0.34	0.28	0.37	0.32	2.37×10^{-4}	1.29 (1.13–1.48)

SALS: sporadic ALS(amyotrophic lateral sclerosis), Chr: chromosome, SNP: single nucleotide polymorphism, MAF: minor allele frequency, OR: odds ratio.

sample compared to an out-bred population up to a distance of 1 Mb. Although the magnitude of this difference appears modest (mean 5.5%) when compared to a true population isolate such as Kosrae, the observed homogeneity operates on several synergistic levels to increase the power of our study to detect genetic variants associated with altered disease risk. First, we observed more tagging pairs at distances below 280 kb. Thus, for a disease-causing variant which has not been directly genotyped, the probability that it will be efficiently tagged by a neighbouring variant is increased (25). Second, in the presence of allelic heterogeneity (a number of SNPs within a given gene each affecting disease susceptibility), homogeneity serves to lower the number of polymorphic loci across that gene, thereby increasing statistical power for detection of the remaining variants (17). Finally, the increase in long-range and background LD overcomes the sensitivity of the case-control paradigm to population sub-structure, owing to the similar and homogenous genetic background of all individuals. These variables are predicted to operate in an independent and multiplicative manner to increase power (17).

Our association study identified 35 SNPs associated with disease within the Irish cohort of 221 ALS cases and 211 controls at an allelic *P*-value less than 0.0001. Consistent with preceding genome scans (14–16), no single locus exceeded the Bonferroni threshold for multiple testing (*P*-value below 1.004×10^{-7}). Previous experience has shown that the top associated SNPs in a genome wide association study that do not clearly exceed Bonferroni most likely represent spurious associations, especially in studies involving relatively small number of samples as was the case in our study. Indeed, none of the 35 most associated SNPs in the Irish cohort were top hits in any of the three previously published genome wide association studies for ALS, nor did the top hits of these studies achieve striking statistical significance in the Irish cohort (14–16). For example, rs2306677, which lies in the gene *ITPR2*, was reported to be associated with ALS in the Dutch ALS population with an allelic *P*-value of 0.005 (MAF cases 0.11, MAF controls 0.06). The same SNP had an allelic *P*-value of 0.64 in the Irish cohort (MAF cases=0.10, MAF controls=0.11). Similarly, Dunckley *et al*. (14) identified association of rs6700125 in the gene *FLJ10986* (*P*-value = 6.0×10^{-4}, MAF cases 0.31, MAF controls 0.39) using a DNA pooling method involving 750 US cases and 766 controls, but again this SNP was not associated with ALS in the Irish cohort (MAF cases = 0.33, MAF controls =0 .34, allelic p for rs6700125 = 0.79). We conclude that discussion of the biological significance of the 35 SNPs that were most associated with ALS in the Irish cohort would be premature.

Instead, the ideal approach to elucidating SNPs that truly alter susceptibility to disease would be to verify findings in separate sets of ALS samples and controls. Such an approach is facilitated by the public availability of genome wide association genotypic data from a cohort of US ALS cases and controls (15) and the availability of allelic results from a previous genome wide association study of Dutch ALS cases (16). The power in each of these individual studies is low to identify loci of moderate effect sizes. To further gene discovery from these efforts, this study performed a joint analysis of a homogeneous Irish cohort and the two previously published genome-wide scans. Joint analysis of the combined cohort of 958 patients with ALS and 932 controls identified seven genomic regions that were associated with disease. rs10260404, an intronic variant in *dipeptidylpeptidase 6*(*DPP6*) on chromosome 7, was the top-ranking SNP in this analysis. The same allele was associated in all three datasets (Irish, $P = 0.029$; US $P = 0.002$; The Netherlands $P = 0.005$). DPP6 (or DPPX) is a transmembrane protein which binds A-type neuronal potassium channels to alter their expression and biophysical properties (30). Interestingly, rs10260404 has also been identified as being significantly associated with disease in a study that compared the most associated SNPs in the Dutch ALS population with the most associated SNPs in the US population (31). It is important to note that the Dutch and the US datasets utilized by van Es *et al*. (31). are the same as the ones presented in this paper. Thus, the new data that is being provided here is replication of the *DPP6* locus using genome-wide data in a cohort of Irish ALS cases and controls. Pooled analysis suggests that possession of the associated allele within *DPP6* increases risk for ALS by 37%.

A second noteworthy finding was two unlinked SNPs (rs6851312 and rs2251316, inter-marker $r^2 = 0.05$) mapping to within 330 kb of each other in the region of chromosome 4 which contains the gene *FBXO8*. The FBXO8 product contributes to phosphorylation-dependent ubiquitination and vesicular protein transport (32), both of which are processes previously implicated as putative mechanisms of motor neuron degeneration (6,33).

We would advocate considerable caution when interpreting the candidate loci discussed. Experience gained through large, multi-centre whole genome studies for common diseases indicates that variants contributing susceptibility may increase risk by as little as 15%. The advantages of our approach are identification of ALS associated SNPs in a genetically homogenous population, followed by replication using full genome-wide data from more out-bred populations. Notwithstanding these benefits, the obvious limitation is the relatively small sample sizes of the three included datasets. Even with the combined set of nearly 1000 cases and 1000 controls, the power to detect a SNP with an odds ratio of 1.5 is only 76% (based on a SNP with MAF of 0.26 and 300 000 markers) (34). The ultimate strategy to isolate the genetic causes of sporadic ALS will be a collaborative approach pooling genome-wide results from several thousand individuals (9). Towards this goal, we have made the raw genotyping sample level data from our study publicly available for download, contributing to a growing and powerful worldwide resource to determine the susceptibility loci for ALS.

MATERIALS AND METHODS

Participants

The Irish DNA samples were collected at the ALS clinic in Beaumont Hospital, Dublin. All patients and controls gave written informed consent to participate in the study, and approval was obtained from the Beaumont Hospital research ethical committee. Detailed phenotype and demographic data are recorded on each individual on the Irish ALS Register, an ongoing

prospective, population based surveillance study. Patients are referred to the specialist clinic from all regions of Ireland.

The demographics of the Irish study population are given in Table 1. All included patients fulfil the criteria for probable or definite sporadic ALS, according to the El Escorial criteria (35), and have been phenotyped by a neurologist with expertise in ALS. Those with familial ALS, based on detailed family history, or with atypical phenotypes, have been excluded. Control individuals with no personal or family history of neurological disease were matched to patients for age and sex. The control samples were collected from unrelated individuals, either spouses of patients with ALS or those accompanying non-ALS patients to neurology clinics. All included participants self-reported Irish Caucasian ethnicity for at least three generations.

Demography of the publicly-available replication series are also given in Table 1. The United States series comprised 276 unrelated, white, non-Hispanic individuals diagnosed with ALS and 271 neurologically normal individuals of similar ethnicity drawn from across the USA. Genotyping was undertaken using Illumina HumanHap 550K SNP chips and full phenotype and raw genotyping data are available from https://queue.coriell.org/Q/index.asp. The Netherlands series included 911 unrelated individuals with all four grandparents born in the Netherlands (461 ALS cases and 450 neurologically normal controls). This cohort was genotyped using the Illumina HumanHap 300K platform. Minor allele frequencies and allelic *P*-values for each genotyped SNP have been made available at http://www.alscentrum.nl/index.php?id=GWA.

Procedures

DNA was extracted from peripheral blood according to standard procedures. Using Illumina Infinium II 550K SNP chips, the Irish samples were assayed for 561 466 SNPs selected from the HapMap Project. Genotyping was performed at the Laboratory of Neurogenetics, Bethesda, according to manufacturer protocols. All samples were genotyped individually. For quality control, chips yielding a call rate below 97.5% were excluded from the study. Genotyping of two replicate samples gave a concordance rate of over 99.9%.

Statistical analysis

To assess differences between the LD structure and homogeneity of the Irish population and an out-bred population of European ancestry, we compared 271 randomly-selected individuals from the Irish study population to 271 control individuals from the United States (15,27). LD was mapped using SNPs across all chromosomes. SNPs that were monomorphic, had a MAF below 0.05 or call rate below 98% in either population were excluded. This process ensured comparison of precisely the same SNP pairs in the same number of individuals in each population (17,27). To estimate LD, we computed the r^2 and LOD score statistics between each pair of SNPs, using the PLINK data analysis toolset (version 0.99s) and Haploview (version 4.0) (22,36). Inter-marker comparisons with LOD scores below 3 in either population were excluded, to out-rule spurious correlations. We then computed the mean r^2 within increasing inter-marker distance bins of 100 kb, and the number of highly correlated SNPs ($r^2 > 0.8$), within increasing inter-marker distance bins of 40 kb.

Prior to association testing, we assessed the Irish cohort for the presence of population structure and cryptic relatedness, both of which may lead to spurious disease associations. Cryptically-related individuals were identified utilizing the identity-by-state (IBS) and pairwise identity-by-decent algorithms as implemented in the PLINK toolset. We then pseudo-randomly selected 2050 unlinked autosomal SNPs and ran the program STRUCTURE version 2.0 assuming from one to six populations to detect the presence of population stratification (23). Genotype data from unrelated individuals from the CEU ($n = 60$), CHB ($n = 45$), JPT ($n = 45$) and YRI ($n = 60$) populations of the HapMap Project (10) were included to facilitate identification of ethnically mismatched individuals. The Irish cohort was also analysed using the program Eigenstrat, which utilizes principal component analysis to detect significant population stratification (24).

For each SNP, we used PLINK to compute allelic χ^2-test association statistics. Deviation from HWE was also computed using the χ^2-test. For replication, we identified SNPs associated with ALS in the Irish, US and Netherlands datasets. Replication was defined as an allelic *P*-value below 0.05 for the same allele for any of the 287 522 SNPs genotyped in all three datasets. Based on this three-study joint analysis paradigm, we estimate that four or five SNPs will be associated for the same allele at a *P*-value below 0.05 by chance alone ($0.025 \times 0.025 \times 0.025 \times 287\ 522$). As allele frequencies, rather than individual genotyping data, were available from The Netherlands study, complete genotyping was assumed for calculation of pooled *P*-values and odds ratios (the mean per SNP call rate was 99.5% in the van ES study) (16). Pooled *P*-values and odds ratios were computed using standard χ^2-testing.

Role of the funding source

The study sponsors had no role in the design, analysis or interpretation of this study.

Full genotyping data

Genotyping data for the Irish cohort may be accessed via the link on the journal website.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We would like to thank the patients, their families and referring neurologists. This study used data from the SNP Resource at the NINDS Human Genetics Resource Center DNA and Cell Line Repository (http://ccr.coriell.org/ninds/). The original genotyping was performed in the laboratory of Drs Singleton and Hardy (NIA, LNG), Bethesda, MD, USA. Access to the samples and these data is in the Coriell website.

Conflict of Interest statement. None declared.

FUNDING

This work was funded by the Muscular Dystrophy Association (United States) (SC, DGB, OH), by the Irish Institute of Clinical Neuroscience Travel Award (SC), by the ALS Association and the Packard Center for ALS Research at Johns Hopkins and by the intramural programmes of the National Institute on Aging (SB, JHD, JC, NW, DGH) and the National Institute on Neurological Disorders and Stroke (BJT).

REFERENCES

1. Lacomblez, L., Bensimon, G., Leigh, P.N., Guillet, P. and Meininger, V. (1996) Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. *Lancet*, **347**, 1425–1431.
2. Logroscino, G., Beghi, E., Zoccolella, S., Palagano, R., Fraddosio, A., Simone, I.L., Lamberti, P., Lepore, V., Serlenga, L. and Registry, S.L.A.P. (2005) Incidence of amyotrophic lateral sclerosis in southern Italy: a population based study. *J. Neurol. Neurosurg. Psychiatry*, **76**, 1094–1098.
3. Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X. *et al.* (1993) Mutations in Cu/Znsuperoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature*, **362**, 59–62.
4. Puls, I., Jonnakuty, C., LaMonte, B.H., Holzbaur, E.L., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J. *et al.* (2003) Mutant dynactin in motorneuron disease. *Nat. Genet.*, **33**, 455–456.
5. Yang, Y., Hentati, A., Deng, H.X., Dabbagh, O., Sasaki, T., Hirano, M., Hung, W.Y., Ouahchi, K., Yan, J., Azim, A.C. *et al.* (2001) 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.*, **29**, 160–165.
6. Nishimura, A.L., Mitne-Neto, M., Silva, H.C., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R., Gillingwater, T., Webb, J. *et al.* (2004) A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am. J. Hum. Genet.*, **75**, 822–831.
7. Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A. *et al.* (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis (ALS4). *Am. J. Hum. Genet.*, **74**, 1128–1135.
8. Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J. *et al.* (2006) ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat. Genet.*, **38**, 411–413.
9. Schymick, J.C., Talbot, K. and Traynor, B.J. (2007) Genetics of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.*, First published Online Oct. 15, PMID 17911166.
10. The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature*, **437**, 1299–1320.
11. McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R. *et al.* (2007) A common allele on chromosome 9 associated with coronary heart disease. *Science*, **316**, 1488–1491.
12. Frayling, T.M. (2007) Genome-wide association studies provide new insights into type 2 diabetes aetiology. *Nat. Rev. Genet.*, **8**, 657–662.
13. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.
14. Dunckley, T., Huentelman, M.J., Craig, D.W., Pearson, J.V., Szelinger, S., Joshipura, K., Halperin, R.F., Stamper, C., Jensen, K.R., Letizia, D. *et al.* (2007) Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N. Engl. J. Med.*, **357**, 775–788.
15. Schymick, J.C., Scholz, S.W., Fung, H.C., Britton, A., Arepalli, S., Gibbs, J.R., Lombardo, F., Matarin, M., Kasperaviciute, D., Hernandez, D.G. *et al.* (2007) Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.*, **6**, 322–328.
16. van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Andersen, P.M., Van Den Bosch, L., de Jong, S.W., van 't Slot, R. *et al.* (2007) ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet Neurol.*, **6**, 869–877.
17. Shifman, S. and Darvasi, A. (2001) The value of isolated populations. *Nat. Genet.*, **28**, 309–310.
18. Wright, A.F., Carothers, A.D. and Piratsu, M. (1999) Population choice in mapping genes for complex diseases. *Nat. Genet.*, **23**, 397–404.
19. Hill, E.W., Jobling, M.A. and Bradley, D.G. (2000) Y chromosome variation and Irish origins. *Nature*, **404**, 351–352.
20. Kelly, F. (1988) *A Guide to Early Irish Law*. Dublin Institute for Advanced Studies, Dublin.
21. McEvoy, B., Richards, M., Forster, P. and Bradley, D.G. (2004) The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic façade of Europe. *Am. J. Hum. Genet.*, **75**, 693–702.
22. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
23. Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
24. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
25. Bonnen, P.E., Pe'er, I., Plenge, R.M., Salit, J., Lowe, J.K., Shapero, M.H., Lifton, R.P., Breslow, J.L., Daly, M.J., Reich, D.E. *et al.* (2006) Evaluating the potential of whole genome studies in Kosrae, an isolated population in Micronesia. *Nat. Genet.*, **38**, 214–217.
26. Zavattari, P., Deidda, E., Whalen, M., Lampis, R., Mulargia, A., Loddo, M., Eaves, I., Mastio, G., Todd, J.A. and Cucca, F. (2000) Major factors influencing linkage disequilibrium by analysis of different chromosome regions in distinct populations: demography, chromosome recombination frequency and selection. *Hum. Mol. Gen.*, **9**, 2947–2957.
27. Tenesa, A., Wright, A.F., Knott, S.A., Carothers, C., Hayward, A., Angius, A., Persico, I., Maestrale, G., Hastie, N.D., Pirastu, M. *et al.* (2004) Extent of linkage disequilibrium in a Sardinian sub-isolate: sampling and methodological considerations. *Hum. Mol. Genet.*, **13**, 25–33.
28. Varilo, T., Paunio, T., Parker, A., Perola, M., Meyer, J., Terwilliger, J.D. and Peltonen, L. (2003) The interval of linkage disequilibrium (LD) detected with microsatellite SNP markers in chromosomes of Finnish populations with different histories. *Hum. Mol. Genet.*, **12**, 51–59.
29. Beghi, E., Logroscino, G., Chio, A., Hardiman, O., Mitchell, D., Swingler, R. and Traynor, B.J. and E.U.R.A.L.S. Consortium (2006) The epidemiology of ALS and the role of population-based registries. *Biochim. Biophys. Acta*, **1762**, 1150–1157.
30. Nadal, M.S., Ozaita, A., Amarillo, Y., Vega-Saenz de Miera, E., Ma, Y., Mo, W., Goldberg, E.M., Misumi, Y., Ikehara, Y., Neubert, T.A. *et al.* (2003) The CD26-related dipeptidylaminopeptidase-like protein DPPX is a critical component of neuronal A-type potassium channels. *Neuron*, **37**, 449–461.
31. van Es, M.A., van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Van den Bosch, L., de Jong, S.W., de Jong, V., Baas, F., van't Slot, R. *et al.* (2008) Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.*, **40**, 29–31.
32. Ilyin, G.P., Rialland, M., Pigeon, C. and Guguen-Guillouzo, C. (2000) cDNA cloning and expression analysis of new members of the mammalian F-box protein family. *Genomics*, **67**, 40–47.
33. Bruijn, L.I., Miller, T.M. and Cleveland, D.W. (2004) Unravelling the mechanisms involved in motor neuron degeneration in ALS. *Annu. Rev. Neurosci.*, **27**, 723–749.
34. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209–213. Genetic power calculator available at http://www.sph.umich.edu/csg/abecasis/CaTS/index.html
35. Brooks, B.R. (1994) El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. *J. Neurol. Sci.*, **124**(suppl), 96–107.
36. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, **21**, 263–265.