European Journal of Human Genetics (2013) 21, 1429–1435

www.nature.com/ejhg

ARTICLE

Homozygosity analysis in amyotrophic lateral sclerosis

Kin Mok[1], Hannu Laaksovirta[2,3], Pentti J Tienari[3], Terhi Peuralinna[3], Liisa Myllykangas[4], Adriano Chiò[5], Bryan J Traynor[3], Michael A Nalls[3], Nicole Gurunlian[1], Aleksey Shatunov[6], Gabriella Restagno[7], Gabriele Mora[8], P Nigel Leigh[9], Chris E Shaw[6], Karen E Morrison[10], Pamela J Shaw[11], Ammar Al-Chalabi[6], John Hardy[1] and Richard W Orrell*[,1]

Amyotrophic lateral sclerosis (ALS) may appear to be familial or sporadic, with recognised dominant and recessive inheritance in a proportion of cases. Sporadic ALS may be caused by rare homozygous recessive mutations. We studied patients and controls from the UK and a multinational pooled analysis of GWAS data on homozygosity in ALS to determine any potential recessive variant leading to the disease. Six-hundred and twenty ALS and 5169 controls were studied in the UK cohort. A total of 7646 homozygosity segments with length >2 Mb were identified, and 3568 rare segments remained after filtering 'common' segments. The mean total of the autosomal genome with homozygosity segments was longer in ALS than in controls (unfiltered segments, $P=0.05$). Two-thousand and seventeen ALS and 6918 controls were studied in the pooled analysis. There were more regions of homozygosity segments per case ($P=1\times10^{-5}$), a greater proportion of cases harboured homozygosity ($P=2\times10^{-5}$), a longer average length of segment ($P=1\times10^{-5}$), a longer total genome coverage ($P=1\times10^{-5}$), and a higher rate of these segments overlapped with RefSeq gene regions ($P=1\times10^{-5}$), in ALS patients than controls. Positive associations were found in three regions. The most significant was in the chromosome 21 SOD1 region, and also chromosome 1 2.9–4.8 Mb, and chromosome 5 in the 65 Mb region. There are more than twenty potential genes in these regions. These findings point to further possible rare recessive genetic causes of ALS, which are not identified as common variants in GWAS.
European Journal of Human Genetics (2013) **21,** 1429–1435; doi:10.1038/ejhg.2013.59; published online 24 April 2013

Keywords: amyotrophic lateral sclerosis; homozygosity; recessive

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease affecting upper and lower motor neurons. It is characterised by rapidly progressive weakness and ultimately death usually from respiratory failure. The majority of cases are idiopathic with about 5% having a familial history.[1] Dominant mutations have been identified in *SOD1*,[2] *TDP43*,[3] *FUS*,[4] *VCP*,[5] and *C9ORF72*.[6,7] Recessive mutations are also known to cause ALS, including the *SOD1* D90A mutation[8–10] and *OPTN*.[11] About 1.6% of the European genome is made up of long segments of homozygosity, representing common ancestral origin.[12] Sporadic ALS may be caused by rare homozygous recessive mutations through inheriting the same defective haplotype. High throughput microarray designed for detecting common single-nucleotide polymorphisms (SNPs) can be used to tag variants and loci. The tag can be extended to haplotypes, including less common ones through combination of SNPs. By running haplotype or homozygosity mapping analysis on these SNPs array data, haplotypes and homozygous segments that are associated with ALS can be identified and used to uncover rare recessive mutations.

We studied patients and controls from the UK and a multinational pooled analysis of GWAS data on homozygosity in ALS to determine any potential regions that may carry recessive disease-associated variants leading to the disease.

METHODS

Patients with ALS fulfilling the El Escorial criteria[13] were recruited from neurology centres in UK. All samples had the SNP genotyped at the UCL Genomics Microarray Centre, using Illumina Human610-Quad BeadChip (Illumina, San Diego, CA, USA) and data assembled in Genome Studio (Illumina). Control data were drawn from the Wellcome Trust Case Control Consortium (WTCCC) data set.[14] The UK ALS sample was as previously published.[15] The other four cohorts were data sets from Finland,[16] Ireland,[17] USA,[18] and Italy.[19] They were all genotyped using Illumina SNP arrays (Table 1).

Sample and SNPs quality control

Standard quality control procedures were applied to each data set before combining. In brief, samples were excluded in PLINK[20] if they had call rates <95%, gender discordance, demonstrated cryptic relatedness (defined as pi_hat >12.5%, effectively removing all first or second degree relatives),

[1]Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, and Department of Clinical Neuroscience, UCL Institute of Neurology, Queen Square, London, UK; [2]Helsinki University Central Hospital, Department of Neurology, Molecular Neurology Research Program, Biomedicum, University of Helsinki, Helsinki, Finland; [3]Molecular Genetics Section and Neuromuscular Diseases Research Group, Laboratory of Neurogenetics, National Institute on Aging, NIH, Bethesda, MD, USA; [4]Department of Pathology, Haartman Institute, University of Helsinki and HUSLAB, and Folkhalsan Institute of Genetics, Helsinki, Finland; [5]Department of Neuroscience, University of Turin and Azienda Ospedaliera Universitaria San Giovanni Battista, Turin, Italy; [6]Medical Research Council Centre for Neurodegeneration Research, King's College London, Institute of Psychiatry, London, UK; [7]Molecular Genetics Laboratory, Azienda Ospedaliera OIRM-Sant'Anna, Turin, Italy; [8]Fondazione Salvatore Mangeri, IRCCS Scientific Institute of Milan, Milan, Italy; [9]Brighton and Sussex Medical School, Trafford Centre for Biomedical Research, University of Sussex, Falmer, UK; [10]School of Clinical and Experimental Medicine, University of Birmingham and Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust, Birmingham, UK; [11]Department of Neuroscience, The Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK
*Correspondence: Dr RW Orrell, Department of Clinical Neuroscience, UCL Institute of Neurology, Rowland Hill Street, London NW3 2PF, UK. Tel: +44 20 7830 2387; Fax: +44 20 7472 6829; E-mail: r.orrell@ucl.ac.uk
Received 13 August 2012; revised 21 February 2013; accepted 28 February 2013; published online 24 April 2013

Table 1 Cohorts in homozygosity analysis

Population	*Case*	*M*	*F*	*Control*	*M*	*F*	*Total*	*Number of SNPs passed QC*	*Illumina array*
Finland	405	199	206	497	104	393	902	318 167	Infinium Human370 and Infinium Human1M BeadChips
Ireland	221	119	102	211	112	99	432	489 106	Infinium II 550 K
Italy	500	269	231	247	135	112	747	500 002	HumanHap550 v1
UK[a]	620	396	224	5169	2608	2561	5789	511 743	Human610-Quad
USA	271	171	100	794	332	462	1065	474 554	Infinium II HumanHap550
Total/merged	2017	1154	863	6918	3291	3627	8935	272 819[b]	

[a]UK case population was previously reported[15] and control taken from WTCCC.
[b]Merged SNPs.

or were outliers from the populations with European ancestry (defined as >3 SD values away from the combined CEU/TSI[21] population mean in component 1 and 2, using PLINK Multi-dimensional scaling plot). SNPs were excluded if they had a minor allele frequency <0.01, Hardy–Weinberg equilibrium ($P<10^{-6}$ in case and $<10^{-4}$ in controls), non-random missing by haplotype ($P<10^{-4}$), or evidence of non-random missing in cases *vs* controls ($P<10^{-4}$). For pooled-cohort analysis, only SNPs that were genotyped, passed quality control and common to all five cohorts were merged.

Runs of homozygosity

Runs of homozygosity segments were called with PLINK v1.07. Default values in PLINK were used for the window scanning of the genome with two exceptions. The window was defined as 50 SNPs (which is the default value), with no heterozygous genotype allowed (default value = 1) and only 1 missing genotype (default value = 5). The more stringent criteria aimed at minimising false positives as genotyping error in Illumina is documented to be much <1%.[22] Homozygosity segments were defined as homozygous segments >2 Mb, taking into account the final pooled five-population cohort had only 273 k SNPs. The use of a longer segment length was aimed at preferentially recruiting more phylogenetically recent segments. The assumption is that pathogenic mutations arise more recently. More recent mutations and their haplotypes are usually longer in length. Analysis was performed in the 22 autosomal chromosomes only. The commands used are given in the Supplementary Method.

Exclusion of common runs of homozygosity segments

Homozygosity segments are more commonly found in regions with lower recombination fraction (the recombination 'cold spot').[23,24] Hundreds of segments could be found. An attempt to eliminate these common runs was made because of our *a priori* assumption of rarity in pathogenic recessive homozygous segments. Common regions were defined as >1% of individuals (cases and controls together) having homozygosity segments of >2 Mb length passing though the region. Kurtosis test (a test for skewness of the data in the sample, R Statiscial Package)[25] was then applied to pick up the segments with skewed length. Rarity was defined as regions with Kurtosis value >10 and segments with length >2 SD from the mean. These segments were retained for association analysis. (An example of pre and post exclusion or filtering is shown in Supplementary Figure 1.)

Burden analysis

Number of segments, proportion of sample with one or more homozygosity segments, total length spanned and average segment size were used for calculating the burden of homozygosity in cases *vs* controls, by adapting the PLINK rare CNV analysis. Means of these parameters were compared using PLINK rare CNV burden analysis (default parameter: 1-sided test; empirical *P*-values generated by permuting case/control status with 100 000 permutations; statistical significance level set as $P<0.05$).

Association analysis

Two types of mapping were used to locate regions in which homozygosity differed between cases and controls. The first was RefSeq gene-based. This gene list was drawn from the list of 19 058 RefSeq genes, based on UCSC browser (built NCBI136/hg18) downloaded from the PLINK website. Number of homozygosity segments overlapping with each RefSeq gene was counted in cases and controls. *P*-values were computed based on 100 000 case/control status permutations adapting from PLINK rare CNV analysis. Correction for genome-wide *P*-value was also generated by PLINK permutation analysis. A second method was based on consensus overlapping regions. The regions were defined in PLINK by lining up all the segments. Each consensus region represents a unique combination of overlapping segments. As in the gene-based method, the number of homozygosity segments overlapping with each region was counted in case and control and the *P*-value derived with permutation. To be more conservative, overall significance was taken as $P<0.01$ in the genome-wide analysis. Annotation was obtained online from NCBI gene.[26]

RESULTS

UK cohort analysis

There was a total of 620 cases (396 males, 224 females), and 5169 controls (2608 males, 2561 females). The cohort consisted of 511 743 SNPs after quality control (Table 1). A total of 7646 homozygosity segments with length >2 Mb (817 in case, 6829 in control) was called with PLINK. Of these, 3568 rare segments remained after filtering out 'common' segments as defined above (388 in case, 3180 in control). There was no significant difference in the mean number of segments per individual (1.318 homozygosity segments per case *vs* 1.321 homozygosity segments per control, $P=0.53$). Mean total coverage of the autosomal genome by homozygosity segments in cases was longer than in controls, 5453 *vs* 5038 kb ($P=0.05$). A greater difference in genome coverage and greater statistical significance was found if the definition of homozygosity was restricted to longer segments. This definition effectively excluded more 'common' homozygosity segments and also retained homozygosity segments and haplotypes phylogenetically more recent in origin, which are usually longer (Supplementary Table 1). Comparison of mean length of homozygosity segments also reached statistical significance if the definition of homozygosity was restricted to longer segments. The same trend was observed whether the analysis was performed in all homozygosity segments or performed on rare homozygosity segments only.

There was no significant association in any RefSeq genes or any consensus regions in the studies within the UK cohort alone. There were nine genes with nominal $P<0.01$ but this did not reach

significance with multiple gene correction in permutation analysis (Supplementary Table 1c).

Combined analysis of five cohorts

Burden of homozygosity. There was a total of 2017 cases and 6918 controls in the pooled five-cohort analysis. The population distribution is shown in Table 1. The final merged cohort consisted of 272 819 SNPs in common. There was a total of 4330 homozygosity segments >2 Mb in ALS cases and 9390 homozygosity segments in controls detected in this pooled five-population cohort. There were more segments per case (2.15 homozygosity segments *vs* 1.36, $P=1\times10^{-5}$), a greater proportion of cases harboured homozygosity (0.71 *vs* 0.66, $P=2\times10^{-5}$), longer average length of homozygosity segments (3524 *vs* 3029 kb, $P=1\times10^{-5}$), a longer total genome coverage ($P=1\times10^{-5}$), and a higher rate of these segments overlapped with RefSeq gene region ($P=1\times10^{-5}$) (Table 2). A similar result was also found when the common segments were filtered out (Supplementary Table 2). The Finnish population contributed more homozygosity segments per sample than the other cohorts. However, the overall significance was retained without the Finnish population. Despite the smaller sample size, the burden of homozygosity segment was also statistically significant in Irish and Italian samples. (Supplementary Table 3).

Table 2 Burden analysis of homozygosity for five cohorts

Test	*Case*	*Control*	*P-value**
Number of segments	4330	9390	
RATE	2.147	1.357	1E−05
PROP	0.7144	0.6646	2E−05
TOTKB	12580	6962	1E−05
AVGKB	3524	3029	1E−05
GRATE	61.91	34.41	1E−05
GPROP	0.703	0.6525	3E−05

RATE: Number of segments per individual.
PROP: Proportion of sample with one or more defined segments.
TOTKB: Mean total kb length spanned by homozygosity segments per individual.
AVGKB: Average segment size.
GRATE: Number of RefSeg gene regions spanned by homozygous segments.
GPROP: Number of homozygosity segments segments with at least one RefSeg gene covered.
**P*-value with 100 000 case–control status swap permutation.

Gene-based association analysis. Three regions of homozygosity were found associated with ALS. (Table 3). A further list of RefSeq gene areas that showed association with *P*-value of <0.05 is shown in Supplementary Table 4. The most significant genome-wide association, corrected with 100 000 case–control status permutations, was centred on Chr 21 *SOD1* region. This was followed by Chr 1 at position 2.9–4.8 Mb and Chr 5 around the 65 Mb region. The same three loci showed evidence of association with ALS even when common segments were retained in the analysis. No associations were found in *TDP43, ANG, FUS, CHMP2B, VCP,* and *C9orf72* regions.

The Chr 21 association was contributed to mainly by the Finnish cohort, which included 40 familial ALS cases with homozygous D90A mutation. Our study showed 20 cases with homozygosity that

Table 3 RefSeq gene-based association analysis

Chr	*Region*	*Nominal P-value#*	*P-value**	*Start of gene*	*End of gene*	*Segments number in cases*	*Segments number in controls*
(A) Chr 1							
1	ACTRT2	1 × 10e−5	0.0096299	2927905	2929327	9	1
1	PRDM16	1 × 10e−5	0.00213	2975603	3345045	10	1
1	ARHGEF16	1 × 10e−5	0.00213	3361006	3387537	10	1
1	MEGF6	1 × 10e−5	0.0096299	3394365	3517919	9	1
1	TPRG1L	1 × 10e−5	0.0096299	3531415	3536554	9	1
1	WDR8	1 × 10e−5	0.0096299	3537198	3556497	9	1
1	TP73	1 × 10e−5	0.0096299	3558988	3640327	9	1
1	KIAA0495	1 × 10e−5	0.00132	3642407	3653746	9	0
1	CCDC27	1 × 10e−5	0.00132	3658821	3678069	9	0
1	LRRC47	1 × 10e−5	0.00132	3686643	3702928	9	0
1	KIAA0562	1 × 10e−5	0.00132	3721203	3763657	9	0
1	DFFB	1 × 10e−5	0.00132	3763704	3791853	9	0
1	C1orf174	1 × 10e−5	0.00132	3795556	3806709	9	0
1	AJAP1	1 × 10e−5	0.0096299	4614964	4743711	9	1
(B) Chr 5							
5	CENPK	1 × 10e−5	0.00422	64849348	64894751	21	14
5	PPWD1	1 × 10e−5	0.00422	64894886	64919126	21	14
5	TRIM23	1 × 10e−5	0.00422	64921262	64955943	21	14
5	C5orf44	1 × 10e−5	0.00422	64956313	64997710	21	14
5	SGTB	1 × 10e−5	0.00422	64997510	65053697	21	14
5	NLN	1 × 10e−5	0.00143	65053840	65155145	21	13
5	ERBB2IP	1 × 10e−5	0.00216	65258139	65412606	20	12
5	SFRS12	1 × 10e−5	0.001	65475840	65512470	20	11
5	MAST4	1 × 10e−5	0.0063099	65927931	66498848	16	8

Pointwise (#) and genome-wide (*) corrected *P*-values performed in PLINK using 100 000 case/control status permutation.
Pointwise (#) and genome-wide (*) corrected *P*-values performed in PLINK using 100 000 case/control status permutation.

spanned *SOD1*. Eleven of these were from the Finnish study. No association with Chr 21 was found if the Finnish cohort was excluded from the analysis.

The most statistically significant region in Chr 1 was found in 2.9–4.8 Mb (nominal P 1 × 10e −5 with 100 000 permutations, corrected for genome-wide, best $P = 0.0013$) (Table 3A). Homozygosity segments at Chr 1 varied from 2 to 12.7 Mb. In the 5′ and 3′ end of this region, there were 10 homozyygous segments in ALS *vs* 1 in control (Figure 1). The 'single' control was contributed by two controls, one from Finnish and one from UK. Each of these two controls overlapped at one of the two ends, Finnish at 5′ and UK at 3′. Within the central core, there was a section that neither of the two controls overlapped, with nine homozygosity segments from ALS cases. A total of 11 ALS cases from different cohorts contributed to the homozygosity segments in this region, 4 Finnish, 4 Italian, 1 Irish, 1 UK, and 1 USA. No single haplotype spanned over this overlapping region.

A total of 14 genes are located in the Chr 1 2.9–4.8 Mb region showing statistically significant association ($P < 0.01$). These included: *ACTRT2, PRDM16, ARHGEF16, MEGF6, TPRG1L, WDR8, TP73, KIAA0495, CCDC27, LRRC47, KIAA0562, DFFB, C1orf174, AJAP1*. The annotations and functions of the potential genes show diverse functions (Supplementary Table 5A).

On Chr 5, the positive associations were clustered around 65 Mb (Figure 2). From 64.8 to 66.5 Mb, nine genes in this region were shown to be significantly in association with ALS (nominal P 1 × 10e −5 with 100 000 permutation, corrected for genome-wide $P < 0.01$) (Table 3B). Among the ALS cases with homozygosity, nine were from Finnish, three Irish, six Italian, two US and one UK. Annotations are shown in Supplementary Table 5B. As in the Chr 1 region, there were no common homozygosity segments around this region. Both analyses yielded similar findings, with or without filtering common segments.

Consensus region-based association analysis. The analysis showed the same three clusters as in the gene-based analysis. Chr 21 was still the most significant. This was followed by a small segment in Chr 5 65.4–65.7 Mb (Table 4), and then by Chr 1 region and other segments in the Chr 5 cluster (Supplementary Table 5). The Chr 5 65.4–65.7 Mb region is an intergenic area with no Refseq gene. Hence this area was omitted when only gene-based association was considered (Figure 2).

DISCUSSION

Recessive mutations have been shown to cause ALS though these are not common.[8] Known recessive genes include *SOD1* (D90A), *ALSIN*, and *OPTN*. *ALSIN* has been associated with rare juvenile onset ALS cases but not adult forms.[27] In *ALSIN* knockout mice, AMPA receptor trafficking is impaired and leads to neuronal degeneration.[28] Genetic ablation of *ALSIN* in *SOD1* (H46R) mice exacerbated motor dysfunction.[29] Alsin also acts as a modifier in Nox2-dependent endosomal reactive oxygen species production in *SOD1*(G93A) glial cells.[30]

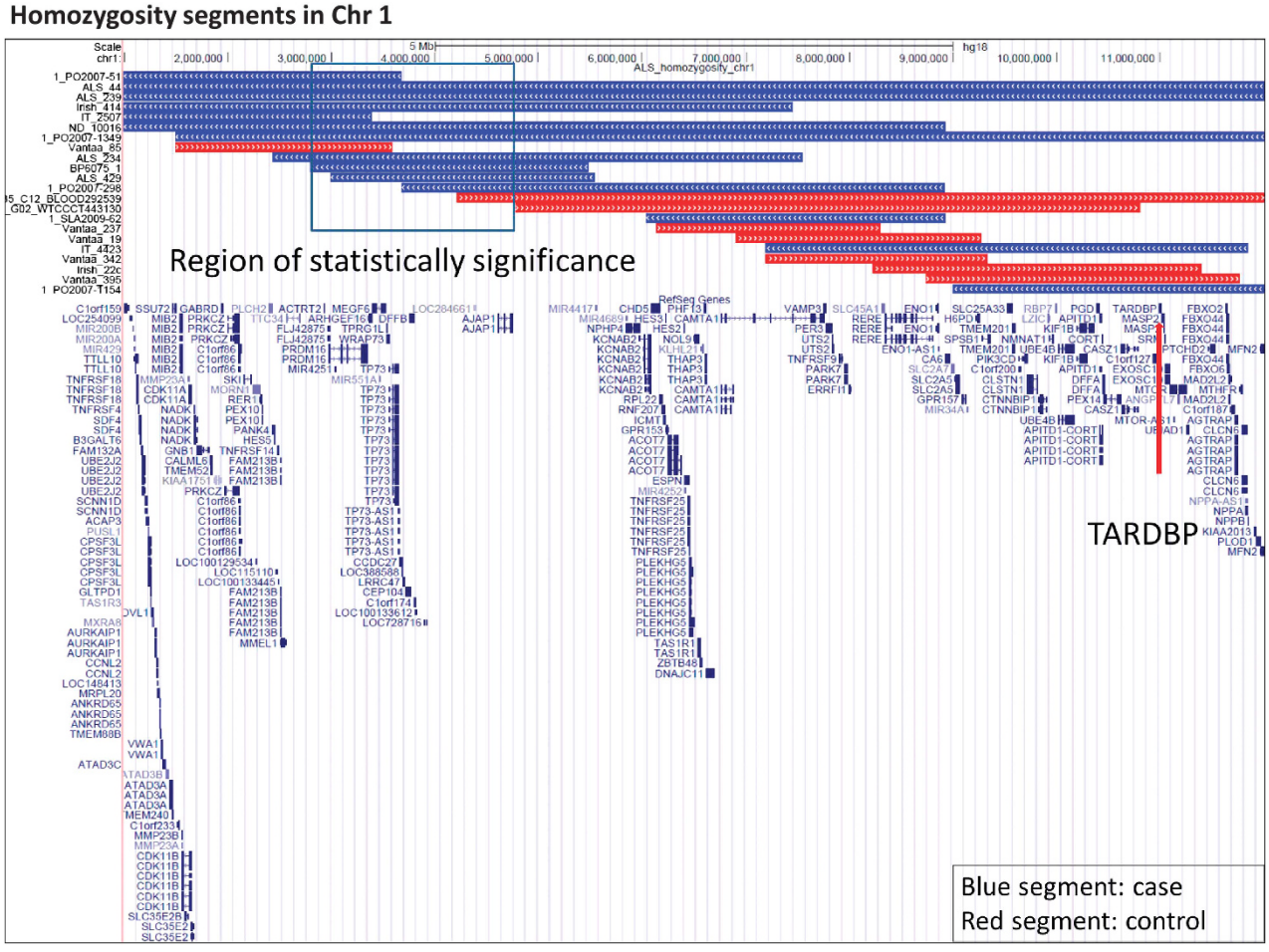

Figure 1 Homozygosity segments in chromosome 1.

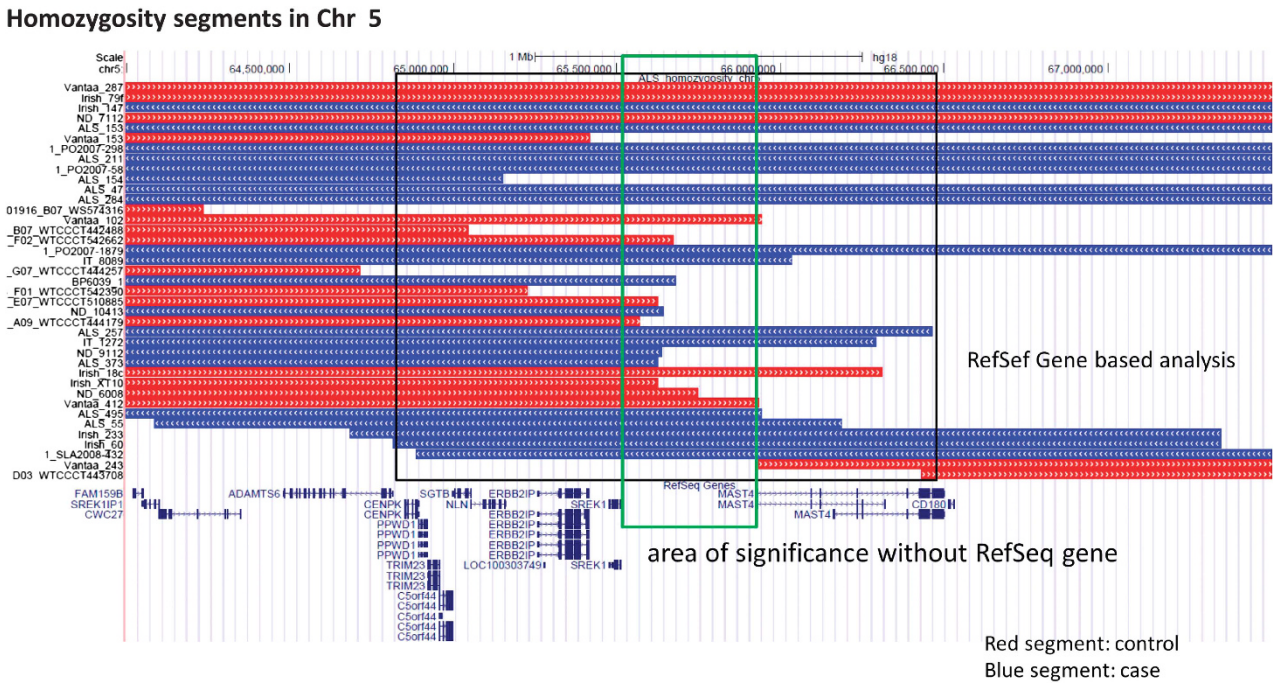

Figure 2 Homozygosity segments in chromosome 5.

Table 4 Chr 5 consensus region analyses with no RefSeq gene

Chr	*Region start*	*Region end*	*Number in case*	*Number in control*	P-*value #*	P-*value**
5	65418257	65569621	20	11	$1 \times 10e-5$	0.0027
5	65569622	65625098	20	10	$1 \times 10e-5$	0.00094
5	65625099	65626254	19	9	$1 \times 10e-5$	0.00096
5	65626255	65639395	19	8	$1 \times 10e-5$	0.00036
5	65639396	65640474	18	8	$2 \times 10e-5$	0.00147
5	65640475	65672811	17	8	$2 \times 10e-5$	0.00386
5	65672812	65678611	17	7	$2 \times 10e-5$	0.00157
5	65678612	65745749	16	7	$2 \times 10e-5$	0.00548

Pointwise (#) and genome-wide (*) corrected *P*-values performed in PLINK using 100 000 case/control status permutation.

Recessive *OPTN* mutations leading to ALS are rarely reported outside Japan, however, optineurin protein was found in inclusions in ALS and other neurodegenerative diseases.[31] Optineurin was also found colocalised with FUS in basophilic inclusions of FUS-linked ALS.[32]

Recessive mutations in ALS are rare. A study of 563 sporadic and 124 familial cases in Caucasians found no non-synonymous mutation.[33] *ALSIN* was not known to cause adult forms of ALS.[27] Nevertheless, recessive mutations may point to important pathological pathways. Overall, we found more homozygosity segments in ALS cases than controls ($P = 1 \times 10^{-5}$), contributed to by Finnish, Irish, and Italian samples. This supports the hypothesis that some ALS cases are due to homozygous recessive loci. Homozygosity generated via identity by descent is known to occur even in outbred populations.[12] Recessive mutants harboured in the haplotype, transmitted to the same individual via this mechanism, could lead to 'sporadic' ALS. Our five-population pooled-cohort analysis shows that not only the number of segments was increased in ALS cases, but the average genome coverage and the mean segment length were also longer in ALS cases. These data suggest a recessive component to the aetiology of sporadic ALS.

The excess homozygosity found on Chr 21, Chr 1, and Chr 5 implies that recessive pathogenic mutants may be located in this region. Detection of the familial *SOD1* mutation contributed by the Finnish cohort indirectly acted as a positive control for our analysis, proving the ability of the method to find pathogenic mutants through significant clustering of homozygosity. Eleven Finnish cases in this study were found to contain homozygous segments >2 Mb at Chr 21 spanning the *SOD1* region. Laaksovirta *et al*[16] found that there were 40 cases of *SOD1* D90A homozygous mutation. Hence, 29 of them were not detected in this study. These remaining 29 undetected cases did not enter into the statistical calculation of association significance. This demonstrates the power of using homozygosity to find a region even with a small number of positive detections. Equally, a small number of false-positive cases will increase the risk of a false-positive association. For this reason that we used strict criteria for preventing false-positive homozygosity call as discussed below.

The discrepancy in Finnish cases is probably due to more than one haplotype harbouring *SOD1* D90A in the Finnish population. Homozygous *SOD1* D90A mutations contributed by different haplotypes would not be read as segments of homozygosity in PLINK. This detection rate still suggests a single dominant haplotype of >2 Mb harbouring the *SOD1* D90A mutation in the Finnish case cohort. It requires a haplotype frequency >0.5 to produce a homozygous haplotype in more than a quarter of the cohort (11/40). Al-Chalabi *et al*[34] suggested that all the *SOD1* D90A recessive mutations had a common founder.

Another explanation for the discrepancy of our homozygosity detection is that the homozygous segments of Laaksovirta *et al*[16] were <2 Mb and thus not detected. Indeed, 22 segments of homozygosity were found spanning *SOD1* if the calling criteria were decreased to 1 Mb. Parton *et al*[35] suggested the disease modifying haplotype ranged

from 97 to 275 kb, well below our 2 Mb segment requirement. We considered the approach of Casey *et al*,[36] using homozygosity of 1 Mb length followed by haplotype analysis. This extracts data from homozygosity and narrows with haplotype comparison to find the target genes. However, our combined cohort has only 273 K SNPs, while Casey *et al* had >1 Mb. This increases the chance of false haplotype call secondary to being identical by state. Also, the cohort of Casey *et al* used family trios for case and control. Our cohort is more diversified. This increases the complexity of clustering and haplotype analysis.

There were 14 genes located in the homozygous region of Chr 1, with diverse functions. The genes act on muscle cell (*PRDM16*), cytoskeleton (*ACTRT2*), RNA function (*KIAA0495*), and apoptosis (*TP73, LRRC47, DFFB*). All of these are potential pathogenic pathways in ALS. Of particular interest is *ARHGEF16* (15 exons). This has a role in protein–protein and protein–lipid interaction. The expression of *ARHGEF16* was recently reported to be upregulated in *CHMP2B* mutant lower motor neuron cells when compared with controls.[37] Another possibility is interaction with *TDP43*. *TDP43* is 7.6 Mb downstream from the 3′ end of ARHGEF16 in the chromosome. It is outside the usual range of cis-control. Chromatin can form loops and put distant segments to adjacent regions for interaction, as for example with beta-globulin.[38] This interaction of homozygous segments in Chr 1 and *TDP-43* could be examined further by examining the control of *TDP43* expression by the segment and vice versa.

There were nine genes located in the associated segment in Chr 5. The length of the homozygosity segment varied from 2.1 to 19 Mb. The best significance was not located in the area of RefSeq genes but in an intergenic region. The molecular genetic pathogenesis of ALS has been proposed to relate to disruption of RNA processing. It is possible that the intergenic region is related to non-coding RNA processes and thus not in the RefSeq gene regions. Based on the Wellcome Trust Rfam database, a ncRNA, small nucleolar RNA U13,[39] was predicted to be in this region.

There are 23 genes in the two regions, and more if the confidence level of association is decreased. This may be examined further by exome sequencing, to determine if there is any common recessive mutation in the regions. Targeted resequencing of the two regions will be required to examine the intergenic regions for mutations other than RefSeq regions.

This type of homozygosity analysis has its own limitations. Given the SNPs in the SNP-array are usually common, there is a high chance that homozygosity of short range, for example, a few SNPs, is the result of being 'identical by state'. There is a balance in retaining more true-positive result at the expense of lots of false-positive results. We took the approach of using more stringent criteria. For example, in the window scanning, we used much harsher criteria than the suggested default value. This approach helps exclude many false positives, but at the expense of excluding true positives. The multiple analyses for the whole pose a similar problem. To prevent an excess of false-positive candidate regions, a high cutoff for the selection will exclude the true rare pathogenic variants. For example, in the pooled-cohorts analysis, a true pathogenic segment that is found three times in cases but not in controls will fail in the statistical test and will be discarded. These analyses usually have a minimum length of 1–2 Mb, generating a large region that rarely gives a single gene for follow-up. Our Chr 21 region (data not shown) showed more than one gene other than *SOD1*. As we were aware that D90A variants were present in the cohorts and D90A is pathogenic, we had no difficulty in identifying the pathogenic gene. This also serves as a positive control in our study. For the other two regions, there is no good evidence to pick one gene over another. Hence, we stress the region as a whole and further examination is required for further dissection, to be followed up by sequencing.

In summary, we found an excess of homozygous segments in a pooled-cohort analysis of GWAS data in sporadic ALS cases. Besides *SOD1*, the best clusters are within Chr 1 around position 3–5 Mb and Chr 5 around position 65–66.5 Mb.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported in part by the Intramural Research Programs of the NIH, the National Institute on Aging (Z01-AG000949-02), and the National Institute of Neurological Disorders and Stroke. Extramural NIH grants R01AG031278 and R01AG038791 supported some family assessments. The research leading to these results has received funding from the European Community's Health Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 259867. We thank the Motor Neurone Disease Association of Great Britain for several grants relating to this work (RWO, AAC, PJS, HM), the ALS Association, The Angel Fund, the ALS Therapy Alliance, and the Wellcome Trust (PJS) for support. This work was also funded by the Reta Lila Weston Foundation, and by an MRC returning scientist (JH) and fellowship (SPB) award, by Microsoft Research Foundation, the ALS Association, Helsinki University Central Hospital, the Finnish Academy, Ministero della Salute, Progetti Finalizzati 2007, Fondazione Vialli e Mauro for ALS, and Federazione Italiana Giuoco Calcio. The authors thank the NIHR specialist Biomedical Research Centre for Mental Health at the South London and Maudsley NHS Foundation Trust (SLaM) and the Institute of Psychiatry, King's College London, and the National Institute for Health Research-funded University College London / University College London Hospitals Biomedical Research Centre.

1 Byrne S, Walsh C, Lynch C *et al*: Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry* 2011; **82**: 623–627.
2 Rosen DR, Siddique T, Patterson D *et al*: Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 1993; **362**: 59–62 (Erratum in: Nature; 364: 362).
3 Sreedharan J, Blair IP, Tripathi VB *et al*: TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* 2008; **319**: 1668–1672.
4 Vance C, Rogelj B, Hortobágyi T *et al*: Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* 2009; **323**: 1208–1211.
5 Johnson JO, Mandrioli J, Benatar M *et al*: Exome sequencing reveals VCP mutations as a cause of familial ALS. *Neuron* 2010; **68**: 857–864 (Erratum in: Neuron 2011;69:397).
6 Renton AE, Majounie E, Waite A *et al*: A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011; **72**: 257–268.
7 DeJesus-Hernandez M, Mackenzie IR, Boeve BF *et al*: Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011; **72**: 245–256.
8 Andersen PM, Nilsson P, Ala-Hurula V *et al*: Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase. *Nat Genet* 1995; **10**: 61–66.
9 Yang Y, Hentati A, Deng HX *et al*: The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. *Nat Genet* 2001; **29**: 160–165.
10 Hadano S, Hand CK, Osuga H *et al*: A gene encoding a putative GTPase regulator is mutated in familial amyotrophic lateral sclerosis 2. *Nat Genet* 2001; **29**: 166–173.
11 Maruyama H, Morino H, Ito H *et al*: Mutations of optineurin in amyotrophic lateral sclerosis. *Nature* 2010; **465**: 223–226.
12 Nalls MA, Simon-Sanchez J, Gibbs JR *et al*: Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. *PLoS Genet* 2009; **5**: e1000415.
13 Brooks BR, Miller RG, Swash M, Munsat TL: World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2000; **1**: 293–299.

14 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven com mon diseases and 3,000 shared controls. *Nature* 2007; **447**: 661–678.
15 Shatunov A, Mok K, Newhouse S *et al*: Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol* 2010; **9**: 986–994 (Erratum in: Lancet Neurol 2011; 10: 205).
16 Laaksovirta H, Peuralinna T, Schymick JC *et al*: Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol* 2010; **9**: 978–985.
17 Cronin S, Berger S, Ding J *et al*: A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum Mol Genet* 2008; **17**: 768–774.
18 Schymick JC, Scholz SW, Fung H *et al*: Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* 2007; **6**: 322–328.
19 Chiò A, Schymick JC, Restagno G *et al*: A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum Mol Genet* 2009; **18**: 1524–1532.
20 Purcell S, Neale B, Todd-Brown K *et al*: PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
21 HapMap, 'HapMapIII Release 2010-05', February 2010; http://hapmap.ncbi.nlm.nih.gov/downloads/genotypes/2010-05_phaseIII/plink_format/.
22 Illumina, 'Genome-Wide DNA Analysis Beadchips' 2010; http://www.illumina.com/Documents/products/datasheets/datasheet_infiniumhd.pdf. http://www.illumina.com/Documents/products/datasheets/datasheet_omni_whole-genome_arrays.pdf. http://www.illumina.com/Documents/products/technotes/technote_intelligent_snp_selection.pdf http://www.solexa.co.uk/downloads/HUMANHAP550_Duo_DataSheet.pdf.
23 Sabatti C, Risch N: Homozygosity and linkage disequilibrium. *Genetics* 2002; **160**: 1707–1719.
24 Wang H, Lin CH, Service S, Chen Y, Freimer N, Sabatti C: Linkage disequilibrium and haplotype homozygosity in population samples genotyped at a high marker density. *Hum Hered* 2006; **62**: 175–189.
25 R Development Core Team. 2008: *R: a language and environment for statistical computing*, in: R Development Core Team (ed.) Vienna, Austria: R Foundation for Statistical Computing, 2008.
26 'NCBI Gene' 2011; http://www.ncbi.nlm.nih.gov/gene.
27 Andersen PM, Al-Chalabi A: Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol* 2011; **11**: 603–615.
28 Lai C, Xie C, McCormack SG *et al*: Amyotrophic lateral sclerosis 2-deficiency leads to neuronal degeneration in amyotrophic lateral sclerosis through altered AMPA receptor trafficking. *J Neurosci* 2006; **26**: 11798–11806.
29 Hadano S, Otomo A, Kunita R *et al*: Loss of ALS2/Alsin exacerbates motor dysfunction in a SOD1-expressing mouse ALS model by disturbing endolysosomal trafficking. *PLoS ONE* 2010; **5**: e9805.
30 Li Q, Spencer NY, Pantazis NJ, JF Engelhardt: Alsin and SOD1(G93A) proteins regulate endosomal reactive oxygen species production by glial cells and proinflammatory pathways responsible for neurotoxicity. *J Biol Chem* 2011; **286**: 40151–40162.
31 Osawa T, Mizuno Y, Fujita Y, Takatama M, Nakazato Y, Okamoto K: Optineurin in neurodegenerative diseases. *Neuropathology* 2011; **31**: 569–574.
32 Ito H, Fujita K, Nakamura M *et al*: Optineurin is co-localized with FUS in basophilic inclusions of ALS with FUS mutation and in basophilic inclusion body disease. *Acta Neuropathol* 2011; **121**: 555–557.
33 Sugihara K, Maruyama H, Kamada M, Morino H, Kawakami H: Screening for OPTN mutations in amyotrophic lateral sclerosis in a mainly Caucasian population. *Neurobiol Aging* 2011; **32**: 1923 e9–10.
34 Al-Chalabi A, Andersen PM, Chioza *et al*: Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor. *Hum Mol Genet* 1998; **7**: 2045–2050.
35 Parton MJ, Broom W, Andersen PM *et al*: D90A-SOD1 mediated amyotrophic lateral sclerosis: a single founder for all cases with evidence for a Cis-acting disease modifier in the recessive haplotype. *Hum Mutat* 2002; **20**: 473.
36 Casey JP, Magalhaes T, Conroy JM *et al*: A novel approach of homozygous haplotype sharing identifies candidate genes in autism spectrum disorder. *Hum Genet* 2012; **131**: 565–579.
37 Cox LE, Ferraiuolo L, Goodall EF *et al*: Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS). *PLoS ONE* 2010; **5**: e9872.
38 Sexton T, Bantignies F, Cavalli G: Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. *Semin Cell Dev Biol* 2009; **20**: 849–855.
39 Wellcome Trust Sanger Institute, 'snoU13' 2011; http://rfam.sanger.ac.uk/family/RF01210.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)