Genome-wide pathway analysis in amyotrophic lateral sclerosis

Y.H. Lee and G.G. Song

Division of Rheumatology, Department of Internal Medicine,
Korea University College of Medicine, Seoul, Korea

Corresponding author: Y.H. Lee
E-mail: lyhcgh@korea.ac.kr

ABSTRACT. The aims of this study were to identify candidate single-nucleotide polymorphisms (SNPs) and mechanisms of amyotrophic lateral sclerosis (ALS) and to generate SNP-to-gene-to-pathway hypotheses. An ALS genome-wide association study (GWAS) dataset that included 483,051 SNPs in 276 patients with ALS and 271 controls of European descent was used in this study. Identify Candidate Causal SNPs and Pathway (ICSNPathway) analysis was applied to the GWAS dataset. ICSNPathway analysis identified 19 candidate SNPs, 8 genes, and 9 pathways, which provided 8 hypothetical biological mechanisms. The strongest hypothetical biological mechanism was that rs9352 alters the role of chromatin assembly factor 1 subunit A in the context of the pathways of chromatin and nucleosome assembly (nominal $P < 0.001$, false discovery rate (FDR) ≤ 0.001, 0.018, respectively). The second strongest was rs1046329 $\rightarrow$ *HILS1* $\rightarrow$ chromatin assembly (nominal $P < 0.001$, FDR = 0.018). The third was rs11100790 $\rightarrow$ *SMARCA5* $\rightarrow$ chromatin and nucleosome assembly (nominal $P < 0.001$, FDR ≤ 0.001, 0.018, respectively). The application of ICSNPathway analysis to the ALS GWAS dataset resulted in the identification of candidate SNPs, pathways, and biological mechanisms that might contribute to ALS susceptibility.

Key words: Amyotrophic lateral sclerosis; Pathway-based analysis; Genome-wide association study

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease of the motor neurons and is characterized by rapidly progressive weakness, muscle atrophy and fasciculations, muscle spasticity, and difficulty in speaking, swallowing, and breathing (Monk and Shaw, 2006). Although the etiology of ALS has not been determined, a genetic component of susceptibility to ALS has been established by case-control, linkage, and familial studies (Schymick et al., 2007b; Chen et al., 2013). A high familial tendency of the disorder and large-effect Mendelian genes have been found to account for 75% of inherited forms (Andersen and Al-Chalabi, 2011; Turner et al., 2013).

Genome-wide association studies (GWASs) offer a powerful means of searching for genes that confer susceptibility to complex diseases (Manolio, 2010). As a result, the number of GWASs being reported is growing rapidly and has led to the discovery and confirmation of new disease genes (Johnson and O'Donnell, 2009). Although large-scale GWASs have been carried out for complex diseases including ALS, much of the genetic component of variations in ALS remains unexplained.

On the basis of the information available, it appears that individual genes and genetic variants make small contributions to the overall risk of disorders by co-interaction . Although genetic signals have been examined at the single-marker level in ALS GWAS studies, the biological mechanisms identified through such analyses remain controversial (Boomsma et al., 2008). Key challenges of GWAS data interpretation are to identify causative single-nucleotide polymorphisms (SNPs) and to provide evidence and hypothetical mechanisms responsible for the traits observed (Wang et al., 2010; Lee and Song, 2012). Thus, we considered that the use of new methods to study existing GWAS data might provide additional biological insights and highlight new candidate genes in ALS. Identify Candidate Causal SNPs and Pathways (ICSNPathway) analysis was developed to identify candidate SNPs and their corresponding candidate pathways using GWAS data by integrating linkage disequilibrium (LD) analysis, functional SNP annotation, and pathway-based analysis (PBA) (Zhang et al., 2011). Accordingly, we applied ICSNPathway analysis to an ALS GWAS dataset to identify candidate causal SNPs and mechanisms of ALS, and to generate SNP-to-gene-to-pathway hypotheses.

MATERIAL AND METHODS

Study population

We used a publicly available ALS GWAS dataset from the National Center for Biotechnology Information (NCBI) dbGap (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000101.v3.p1), which included the genotypes of 512,294 SNPs obtained from an Illumina Infinium II HumanHap550 SNP chip. This dataset comprised information from 276 patients with ALS and 271 control individuals of European descent (Schymick et al., 2007a). Samples were obtained from the National Institute of Neurological Disorders and Stroke (NINDS) Neurogenetics Repository at the Coriell Institute for Medical Research, Camden, NJ, USA. DNA from 276 unique, unrelated, Caucasian, non-Hispanic individuals diagnosed with sporadic ALS was selected for the analysis. The age of the patients at the onset of disease ranged from 26 to 87 years (mean 54.8 years). The panels utilized that contained samples from neurologically normal controls were NDPT002, NDPT006, and NDPT008. The age at sample collection of the control cohort ranged from 55 to 88 years (mean, 68 years).

The dataset was filtered to remove individuals with P < 0.0001 for Hardy-Weinberg violation, and a call rate of <98% was used to reduce the impact of genotyping errors. The 483,051 SNPs analyzed in the study passed the quality control filters.

Identification of candidate causal SNPs and pathways

ICSNPathway analysis occurs in two stages (Zhang et al., 2011). The first involves the pre-selection of candidate causal SNPs by LD analysis and functional SNP annotation based on the most significant SNPs. The second involves the annotation of biological mechanisms to preselected candidate causal SNPs using a PBA algorithm called improved gene-set enrichment analysis (*i*-GSEA). A full list of ALS GWAS SNP P values was entered into the ICSNPathway analysis. LD analysis searches for the most significant SNPs in LD within a GWAS dataset to identify more possible candidate causal SNPs based on an extended data set, which includes HapMap data (International HapMap Consortium et al., 2010). The use of functional SNPs in ICSNPathway analysis preselects candidate causal SNPs based on their function, which is important for understanding the underlying genetics of human health. Functional SNPs are defined as SNPs that may alter protein or gene expression or the role of a protein in the context of a pathway. Functional SNPs include deleterious and non-deleterious nonsynonymous SNPs, and SNPs that cause the gain or loss of a stop codon, result in a frame shift, or are located in essential splice sites or in regulatory regions. The ICSNPathway server applies a PBA algorithm, *i*-GSEA, to the full list of GWAS SNP P values to detect pathways associated with traits. Briefly, the process is as follows: 1) Each SNP is mapped to its nearest gene according to the SNP and gene localization in the Ensembl 61 database (http://www.wnsembl.org/biomart/martview), and the maximum $t = $ -log (P value) values of SNPs mapped to genes are assigned to represent those genes. All genes are subsequently ranked by decreasing representative t values. 2) For each pathway S, ES [enrichment score, i.e., a Kolmogrov-Smirnov-like running-sum statistic with weight (a)] is calculated, which measures the tendency for genes of a pathway to be located at the top of the ranked gene list. 3) The ES is then converted to a significant-proportion-based ES ($SPES$) by multiplying the ES by m_1/m_2, where m_1 is the proportion of significant genes (defined as genes mapped with at least one in the top 5% of the most significant SNPs in the GWAS) for pathways S, and m_2 is the proportion of significant genes of all genes in the GWAS. 4) SNP label permutation and normalization are employed to generate the distribution of $SPES$s and to correct gene variation (the bias caused by different genes with different numbers of mapped SNPs) and pathway variation (the bias due to different pathways with different numbers of genes). 5) Based on the distribution of $SPES$ values generated by permutation, a nominal P value is calculated, and a false discovery rate (FDR) is computed for a multiple testing correction.

The phrase "the most significant SNPs" refers to SNPs with a P value below a certain threshold. This threshold can be specified from the GWAS SNP P values. ICSNPathway analysis was used to examine significant pathways from the original GWAS to choose the P value threshold (<0.01) used in this study. Two parameters were set for the analysis. The first parameter was "within gene", meaning that only P values of SNPs located within genes were utilized in the PBA algorithm. The second parameter was an FDR cutoff (0.05) for multiple testing corrections. Control of the FDR is preferred for large-scale testing. Defined as the expected proportion of false positives among all significant tests, it allows researchers to identify a set of "candidate positives", a high proportion of which are likely to be true positives. The FDR, a permutation-based approach for multiple comparisons, was used to identify statistical-

ly significant genes. There were no specific criteria to select gene numbers. We used minimum and maximum cut-offs of 5 and 100, respectively, to avoid overly narrow or broad functional categories. To avoid stochastic bias and the inclusion of a general biological process, we discarded pathways that contained over 100 genes. Of the several options available for pathway annotation, we chose four pathway databases, namely, the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al., 2010), BioCarta, Gene Ontology (GO) biological process (Ashburner et al., 2000), and GO molecular function, to ensure comprehensive coverage and high-quality information for well-defined pathways.

When a candidate SNP is not present in a particular genotyping array, proxy SNPs in LD with that candidate SNP can be identified as candidate SNPs themselves on the basis of observed LD patterns in HapMap. The extensive nature of LD can confound the interpretation of an association signal as a true causal variant and can lie at a considerable distance from the initial association signal. HapMap provides information about genomic locations, alleles, and LD patterns for a large fraction of common variants in the human genome. The screening for non-acceptable polymorphisms (SNAP) method was developed to identify and annotate nearby SNPs in LD (proxies) by HapMap. In this study, we used SNAP to 1) find proxy SNPs; 2) determine whether SNP proxies were in genes; 3) resolve whether associations from multiple SNPs represent similar associations; 4) plot regional views of associations or LD structures; and 5) retrieve annotations for SNPs (Johnson et al., 2008).

RESULTS

Candidate SNPs and pathways resulting from the ALS GWAS

By utilizing the 483,051 GWAS SNP P values as input and filtering out the most significant SNPs (P < 0.01), ICSNPathway analysis identified 19 candidate SNPs, 8 genes, and 9 pathways (Tables 1-3; Figure 1) associated with ALS.

Table 1. Candidate causal SNPs.

Candidate causal SNP	Functional class	Gene	Candidate causal pathway[a]	$-\log_{10}(P)$[b]	In LD with:	r^2	D'	$-\log_{10}(P)$[c]
rs9352	Nonsynonymouscoding (deleterious)	*CHAF1A*	1, 2	2.100	rs9352	-	-	2.100
rs1046329	Nonsynonymouscoding	*HILS1*	1	-	rs1061947	0.933	1.0	2.691
rs11100790	Stop;gain of function	*SMARCA5*	1, 2	-	rs11730153	0.917	1.0	2.456
rs3127328	Nonsynonymouscoding	*T*	3, 8, 9	-	rs2028791	1.0	1.0	3.394
rs1801046	Nonsynonymouscoding	*C1R*	4	2.048	rs1801046	-	-	2.048
rs12711521	Nonsynonymouscoding (deleterious)	*MASP2*	4	2.309	rs12711521	-	-	2.309
rs3795524	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3795523	Nonsynonymouscoding (deleterious)	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3795522	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3795519	Nonsynonymouscoding (deleterious)	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3795518	Nonsynonymouscoding (deleterious)	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs12067133	Nonsynonymouscoding	*CENPF*	5, 6	1.767	rs17023380	1.0	1.0	2.040
rs3795517	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3790646	Nonsynonymouscoding (deleterious)	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3790647	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3748695	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs12058704	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs3748696	Nonsynonymouscoding	*CENPF*	5, 6	-	rs17023380	1.0	1.0	2.040
rs2525574	Stop;los of functions	*NF1*	7	2.713	rs2525574	-	-	2.713

SNP = single nucleotide polymorphism; LD= linkage disequilibrium. [a]Indices of pathways ranked by significance (false discovery rate). [b]$\log_{10}(P)$ values of candidate causal SNPs in the original genome-wide association studies(GWASs). [c]$\log_{10}(P)$ values of SNPs in LD with candidate causal SNPs in the original GWASs."-" = SNPs not represented in the original GWAS.

Gene	Function
CHAF1A	Core component of the CAF-1 complex, thought to mediate chromatin assembly in DNA replication and repair. Assembles histone octamers onto replicating DNA *in vitro*. CAF-1 performs the first step of the nucleosome assembly process, bringing newly synthesized histones H3 and H4 to replicating DNA; histones H2A/H2B can bind to this chromatin precursor subsequent to DNA replication to complete the histone octamer. CHAF1A binds to histones H3 and H4. It may play a role in heterochromatin maintenance in proliferating cells by bringing newly synthesized cbx proteins to heterochromatic DNA replication loci (by similarity). The CCR4-NOT complex functions as a general transcription regulation complex. Also involved in vitamin D-coupled transcription regulation via its association with the WINAC complex, it is a chromatin-remodeling complex recruited by the vitamin D receptor (VDR), which is required for the ligand-bound VDR-mediated transrepression of the *CYP27B1* gene.
HILS1	DNA-binding protein that may be implicated in chromatin remodeling and/or transcriptional regulation during spermiogenesis, the process of spermatid maturation into spermatozoa.
SMARCA5	Helicase that possesses intrinsic ATP-dependent nucleosome-remodeling activity. Complexes containing SMARCA5 are capable of forming ordered nucleosome arrays on chromatin; this may require intact histone H4 tails. Also required for replication of pericentric heterochromatin in S-phase specifically in conjunction with BAZ1A. Probably plays a role in repression of polI-dependent transcription of the rDNA locus, through the recruitment of the SIN3/HDAC1 corepressor complex to the rDNA promoter. Essential component of the WICH complex, a chromatin-remodeling complex that mobilizes nucleosomes and reconfigures irregular chromatin to a regular nucleosomal array structure. The WICH complex regulates the transcription of various genes, has a role in RNA polymerase I and III transcription, mediates histone H2AX phosphorylation at Tyr-142, and is involved in the maintenance of chromatin structures during DNA replication processes. Essential component of the nucleolar remodeling complex (NoRC), a complex that mediates the silencing of a fraction of rDNA by recruiting histone-modifying enzymes and DNA methyltransferases, leading to heterochromatin formation and transcriptional silencing.
T	Involved in the transcriptional regulation of genes required for mesoderm formation and differentiation. Binds to a palindromic site (called a T site) to activate gene transcription.
C1R	The C1r B chain is a serine protease that combines with C1q and C1s to form C1, the first component of the classical pathway of the complement system.
MASP2	Serum protease that plays an important role in the activation of the complement system via mannose-binding lectin. After activation by autocatalytic cleavage, it cleaves C2 and C4, leading to their activation and to the formation of C3 convertase.
CENPF	Required for kinetochore function and chromosome segregation in mitosis. Required for kinetochore localization of dynein, LIS1, NDE1, and NDEL1. Regulates recycling of the plasma membrane by acting as a link between recycling vesicles and the microtubule network though its association with STX4 and SNAP25. Acts as a potential inhibitor of pocket protein-mediated cellular processes during development by regulating the activity of retinoblastoma (RB) proteins during cell division and proliferation. May play a regulatory or permissive role in the normal embryonic cardiomyocyte cell cycle and in promoting continued mitosis in transformed, abnormally dividing neonatal cardiomyocytes. Interaction with RB directs embryonic stem cells toward a cardiac lineage. Involved in the regulation of DNA synthesis and hence, cell cycle progression, via its C-terminus. Has a potential role in regulation of skeletal myogenesis and in cell differentiation in embryogenesis. Involved in dendritic cell regulation of T-cell immunity against chlamydia.
NF1	Stimulates the GTPase activity of Ras. Shows greater affinity for Ras GAP but lower specific activity. May be a regulator of Ras activity.

The top five candidate SNPs were rs9352 [$-\log_{10}(P) = 2.100$], rs1046329 [$-\log_{10}(P) = 2.691$], rs11100790 [$-\log_{10}(P) = 2.456$], rs3127328 [$-\log_{10}(P) = 3.394$], and rs1801046 [$-\log_{10}(P) = 2.048$]. Of these, two, rs9352 and rs1801046, were not in LD with any SNP. SNP rs1046319, which was not represented in the original GWAS, was in LD with rs1061947 ($r^2 = 0.933$). SNP rs11100790, which was not represented in the original GWAS, was in LD with rs11730153 ($r^2 = 0.917$), and SNP rs3127328, which was not represented in the original GWAS, was in LD with rs2028791 ($r^2 = 1.0$). Biological mechanisms were identified based on the assumption that the candidate SNP might alter the role of its corresponding gene/protein in the context of the pathway(s) associated with certain traits. The 19 candidate SNPs included in 9 candidate pathways indicated 8 hypothetical biological mechanisms. The strongest hypothetical biological mechanism was that rs9352 alters the role of chromatin assembly factor 1 subunit A (CHAF1A) in the context of the pathways of chromatin and nucleosome assembly (nominal P < 0.001, FDRs ≤ 0.001, 0.018, respectively). The second strongest was rs1046329 → *HILS1* → chromatin assembly (nominal P < 0.001, FDR = 0.018). The third was rs11100790 → *SMARCA5* → chromatin and nucleosome assembly (nominal P < 0.001, FDRs ≤ 0.001, 0.018, respectively). The fourth was rs3127328→ *T* → anterior/posterior axis specification, tripartite regional subdivision, and embryonic axis specification (nominal P < 0.001, FDRs = 0.031, 0.047, and 0.047, respectively). The fifth was rs1801046 → *C1R* → complement activation, classical pathway (nominal P < 0.001, FDR = 0.032) (Tables 1-3).

Table 3. Candidate causal pathways.

Index	Candidate causal pathway	Description	Nominal P	FDR
1	Chromatin assembly	GO:0031497. The assembly of DNA, histone proteins, and other associated proteins into chromatin structure, beginning with the formation of the basic unit, the nucleosome, followed by organization of the nucleosomes into higher order structures, ultimately giving rise to a complex organization of specific domains within the nucleus.	<0.001	<0.001
2	Nucleosome assembly	GO:0006334. The aggregation, arrangement, and bonding together of a nucleosome, the bead-like structural units of eukaryotic chromatin composed of histones and DNA.	<0.001	0.018
3	Anterior/posterior axis specification	GO:0009948. The establishment, maintenance, and elaboration of the anterior/posterior axis. The anterior-posterior axis is defined by a line that runs from the head or mouth of an organism.	<0.001	0.031
4	Complement activation, classical pathway	GO:0006958. Any process involved in the activation of any of the steps of the classical pathway of the complement cascade that allows for the direct killing of microbes, disposal of immune complexes, and regulation of other immune processes.	<0.001	0.032
5	Interphase	GO:0051325. Progression through the interphase, the stage of cell cycle between successive rounds of chromosome segregation. Canonically, the interphase is the stage of the cell cycle during which the biochemical and physiological functions of the cell are performed and replication of chromatin occurs.	0.001	0.039
6	Interphase of mitotic cell cycle	GO:0051329. Progression through the interphase, the stage of the cell cycle between successive rounds of mitosis. Canonically, interphase is the stage of the cell cycle during which the biochemical and physiological functions of the cell are performed and replication of chromatin occurs.	0.001	0.042
7	Negative regulation of map kinase activity	GO:0043407. Any process that stops, prevents, or reduces the frequency, rate, or extent of MAP kinase activity.	0.002	0.045
8	Tripartite regional subdivision	GO:0007351. Subdivision of the embryo along the anterior/posterior axis into anterior, posterior, and terminal regions.	<0.001	0.047
9	Embryonic axis specification	GO:0000578. The establishment, maintenance, and elaboration of a pattern along a line or a point in an embryo.	<0.001	0.047

FDR = false discovery rate.
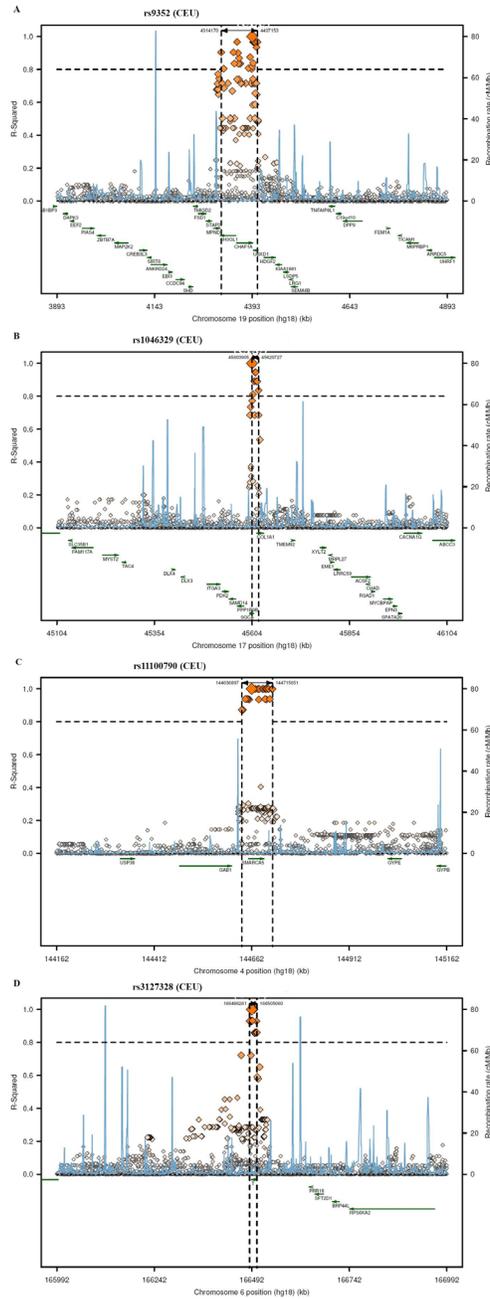
Figure 1. Regional LD plots of the rs9352 (*CHAF1A*) (**A**), rs1046329 (*HILS1*) (**B**), rs11100790 (*SMARCA5*) (**C**), and rs3127328 (*T*) (**D**) SNPs. SNPs are plotted along with their proxies (based on 1000 genomes pilot 1 CEU) as a function of genomic location, annotated by the recombination rate across the locus (light-blue line). On the y-axis, pairwise *r²* values are provided for each proxy SNP using color codes. LD = linkage disequilibrium; SNP = single nucleotide polymorphism; CEU = Northern Europeans from Utah.

Genes without a known immunological function are of considerable interest because they might lead to the identification of novel mechanisms for ALS susceptibility. The analysis in the current study suggested that the pathways and genes found by pathway-based analysis might contribute to ALS susceptibility (Tables 1-3).

DISCUSSION

GWASs have been used successfully to identify novel common genetic variants that contribute to the susceptibility to complex diseases (Manolio, 2010). However, individual GWASs are limited in terms of their ability to identify new loci because a limited set of variants are genotyped and the variant reported as showing disease association is unlikely. Reported loci are those that reach a certain stringent statistical "genome-wide" significance criterion, whereas hundreds of thousands of other genotyped markers receive little attention. Multiple related genes in a pathway might work together to confer disease susceptibility, and some of these genes may not individually reach genome-wide significance in any single GWAS. Thus, PBA is required to identify new loci associated with the susceptibility to complex diseases (Elbers et al., 2009; Zeggini and Ioannidis, 2009).

A complex molecular network and various cellular pathways play key roles in the development of ALS. If a specific pathway was relevant to disease susceptibility, association signals would be expected to be overrepresented for SNPs in that pathway (Fridley and Biernacka, 2011). Given the limited power of GWAS to detect single SNP associations, we adopted a pathway-based approach, because such approaches take into account the biological interplay between genes and provide insights into how multiple genes might contribute to the pathogenesis of diseases (Lee et al., 2012a,b).

In the present study, we identified 19 candidate SNPs, 8 genes, and 9 pathways for ALS susceptibility by using ICSNPathway analysis, which provided 8 hypothetical biological mechanisms. The most significant SNP-to-gene-to-effect hypothesis identified in this study was that rs9352 might alter the function of CHAF1A in the context of the pathways of chromatin and nucleosome assembly. CHAF1A is a core component of the CAF-1 complex, which is thought to mediate chromatin assembly during DNA replication and repair (Kaufman et al., 1995). However, the association between CHAF1A and ALS has not yet been well studied. The second most significant hypothesis was rs1046329 → *HILS1* → chromatin assembly. The histone H1-like protein in spermatids 1 (HILS1) protein is implicated in chromatin remodeling and/or transcriptional regulation during spermiogenesis (Teng et al., 2010). The third most significant was rs6720481 → rs11100790 → *SMARCA5* → chromatin assembly and nucleosome assembly. *SMARCA5* encodes SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5, which has helicase and ATPase activities and is thought to regulate the transcription of certain genes by altering the chromatin structure around those genes (Soldi and Bonaldi, 2013). These findings are novel, and further validation for the observed relationships is needed.

None of the 8 candidate genes have previously been reported in association with ALS; however, our ICSNPathway analysis identified promising genes and pathways that are related to the immune system. Of the 8 candidate genes, 3 are considered to be associated with immune response that might be linked to ALS: *CHAF1A*, *C1R*, and *MASP2*. There is increasing evidence that the development of mental and neurologic disorders, including ALS, is related to complex pathways involving the immune system and inflammatory response (Pardo et al.,

2005; Dantzer et al., 2008). The C1r B chain is a serine protease that combines with C1q and C1s to form C1, the first component of the classical pathway of the complement system (Wijeyewickrema et al., 2013). Mannan-binding lectin serine peptidase 2 (MASP2) plays an important role in the activation of the complement system via mannose-binding lectin (Duncan et al., 2012). After activation by autocatalytic cleavage, MASP2 cleaves C2 and C4, leading to their activation and to the formation of the C3 convertase. High mRNA and protein levels of members of the classical complement pathway have been found in ALS, suggesting that a complement-driven immune response might contribute to the progression of inflammation, resulting in motor neuron injury (Sta et al., 2011). Although the mechanisms of immune-mediated damage in ALS remain unknown, our findings suggest the need for further investigation of the roles of immunity and inflammation in ALS.

Pathway-based approaches, such as ICSNPathway analysis, were developed to solve the challenges presented by GWAS data interpretation (Zhang et al., 2011). However, the incomplete annotation of the human genome is an important limitation of this approach. Furthermore, a proportion of human genes remain uncharacterized, and thus, these genes have not been mapped to predicted pathways. Therefore, results should be interpreted with caution, and additional replication studies are needed to confirm the candidate SNPs and genes underlying the associations between pathways and ALS identified during the present study. Nevertheless, pathway-based approaches play a complementary role to classic methods in the identification of novel genes that confer disease susceptibility. Thus, the results obtained in the present study might lead to the formulation of new hypotheses, which would require additional validation.

In summary, we examined an ALS GWAS dataset to identify genetic associations with ALS at both the SNP and pathway levels. By applying ICSNPathway analysis to this dataset, we identified 19 candidate SNPs, 8 genes (including *CHAF1A*, *C1R*, and *MASP2*), and 9 pathways involving chromatin and nucleosome assembly and complement activation, which might contribute to ALS susceptibility. Further studies are needed to confirm and explore genetic variation in molecular pathways associated with ALS.

Conflicts of interest

The authors declare no conflict of interest.

REFERENCES

Andersen PM and Al-Chalabi A (2011). Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat. Rev. Neurol.* 7: 603-615.

Ashburner M, Ball CA, Blake JA, Botstein D, et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25-29.

Boomsma DI, Willemsen G, Sullivan PF, Heutink P, et al. (2008). Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur. J. Hum. Genet.* 16: 335-342.

Chen S, Sayana P, Zhang X and Le W (2013). Genetics of amyotrophic lateral sclerosis: an update. *Mol. Neurodegener.* 8: 28.

Dantzer R, O'Connor JC, Freund GG, Johnson RW, et al. (2008). From inflammation to sickness and depression: when the immune system subjugates the brain. *Nat. Rev. Neurosci.* 9: 46-56.

Duncan RC, Bergström F, Coetzer TH, Blom AM, et al. (2012). Multiple domains of MASP-2, an initiating complement protease, are required for interaction with its substrate C4. *Mol. Immunol.* 49: 593-600.

Elbers CC, van Eijk KR, Franke L, Mulder F, et al. (2009). Using genome-wide pathway analysis to unravel the etiology

of complex diseases. *Genet. Epidemiol.* 33: 419-431.

Fridley BL and Biernacka JM (2011). Gene set analysis of SNP data: benefits, challenges, and future directions. *Eur. J. Hum. Genet.* 19: 837-843.

International HapMap Consortium, Altshuler DM, Gibbs RA, Peltonen L, et al. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* 467: 52-58.

Johnson AD and O'Donnell CJ (2009). An open access database of genome-wide association results. *BMC Med. Genet.* 10: 6.

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, et al. (2008). SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* 24: 2938-2939.

Kanehisa M, Goto S, Furumichi M, Tanabe M, et al. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.* 38: D355-360.

Kaufman PD, Kobayashi R, Kessler N and Stillman B (1995). The p150 and p60 subunits of chromatin assembly factor I: a molecular link between newly synthesized histones and DNA replication. *Cell* 81: 1105-1114.

Lee YH and Song GG (2012). Pathway analysis of genome-wide association studies on uric acid concentrations. *Hum. Immunol.* 73: 805-810.

Lee YH, Bae SC, Choi SJ, Ji JD, et al. (2012a). Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis. *Mol. Biol. Rep.* 39: 10627-10635.

Lee YH, Choi SJ, Ji JD and Song GG (2012b). Genome-wide pathway analysis of a genome-wide association study on psoriasis and Behcet's disease. *Mol. Biol. Rep.* 39: 5953-5959.

Manolio TA (2010). Genomewide association studies and assessment of the risk of disease. *N. Engl. J. Med.* 363: 166-176.

Monk PN and Shaw PJ (2006). ALS: life and death in a bad neighborhood. *Nat. Med.* 12: 885-887.

Pardo CA, Vargas DL and Zimmerman AW (2005). Immunity, neuroglia and neuroinflammation in autism. *Int. Rev. Psychiatry* 17: 485-495.

Schymick JC, Scholz SW, Fung HC, Britton A, et al. (2007a). Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 6: 322-328.

Schymick JC, Talbot K and Traynor BJ (2007b). Genetics of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 16: R233-242.

Soldi M and Bonaldi T (2013). The proteomic investigation of chromatin functional domains reveals novel synergisms among distinct heterochromatin components. *Mol. Cell. Proteomics* 12: 764-780.

Sta M, Sylva-Steenland RM, Casula M, de Jong JM, et al. (2011). Innate and adaptive immunity in amyotrophic lateral sclerosis: evidence of complement activation. *Neurobiol. Dis.* 42: 211-220.

Teng YN, Kuo PL, Cheng TC and Liao MH (2010). Histone gene expression profile during spermatogenesis. *Fertil. Steril.* 93: 2447-2449.

Turner MR, Goldacre R, Ramagopalan S, Talbot K, et al. (2013). Autoimmune disease preceding amyotrophic lateral sclerosis: an epidemiologic study. *Neurology* 81: 1222-1225.

Wang K, Li M and Hakonarson H (2010). Analysing biological pathways in genome-wide association studies. *Nat. Rev. Genet.* 11: 843-854.

Wijeyewickrema LC, Yongqing T, Tran TP, Thompson PE, et al. (2013). Molecular determinants of the substrate specificity of the complement-initiating protease, C1r. *J. Biol. Chem.* 288: 15571-15580.

Zeggini E and Ioannidis JP (2009). Meta-analysis in genome-wide association studies. *Pharmacogenomics* 10: 191-201.

Zhang K, Chang S, Cui S, Guo L, et al. (2011). ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. *Nucleic Acids Res.* 39: W437-W443.