A multi-ethnic meta-analysis identifies novel genes, including ACSL5, associated with amyotrophic lateral sclerosis

Ryoichi Nakamura et al.

Amyotrophic lateral sclerosis (ALS) is a devastating progressive motor neuron disease that affects people of all ethnicities. Approximately 90% of ALS cases are sporadic and thought to have multifactorial pathogenesis. To understand the genetics of sporadic ALS, we conducted a genome-wide association study using 1,173 sporadic ALS cases and 8,925 controls in a Japanese population. A combined meta-analysis of our Japanese cohort with individuals of European ancestry revealed a significant association at the ACSL5 locus (top SNP \( p = 2.97 \times 10^{-8} \)). We validated the association with ACSL5 in a replication study with a Chinese population and an independent Japanese population (1941 ALS cases, 3821 controls; top SNP \( p = 1.82 \times 10^{-4} \)). In the combined meta-analysis, the intronic ACSL5 SNP rs3736947 showed the strongest association (\( p = 7.81 \times 10^{-11} \)). Using a gene-based analysis of the full multi-ethnic dataset, we uncovered additional genes significantly associated with ALS: ERGIC1, RAPGEF5, FNBP1, and ATXN3. These results advance our understanding of the genetic basis of sporadic ALS.

#A list of authors and their affiliations appears at the end of the paper.
myotrophic lateral sclerosis (ALS) is the most common motor neuron disease characterised by progressive skeletal muscle atrophies that lead to death mostly within 3–5 years from disease onset. Mutations in more than 30 genes have been identified to cause ALS, including SOD1, TARDBP, FUS, and C9orf72. Approximately 50–70% of ALS patients with a family history of the disease can be attributed to mutations in these genes; however, these mutations are found in only 3% of the Japanese patients with sporadic ALS. A twin study estimated the heritability of sporadic ALS to be 61%\(^2\); therefore, sporadic ALS is thought to be a multifactorial disease to which multiple genetic and environmental factors contribute. Previous genome-wide association studies (GWAS) reported six common loci showing genome-wide significant associations with ALS, which only explained 0.2% of its heritability\(^4\). To elucidate the pathophysiology of sporadic ALS and to develop appropriate therapies, the genetic background of sporadic ALS needs to be understood more clearly.

Assuming that common causal variants in different populations are important in investigating the pathophysiology of sporadic ALS, cross-ethnic meta-analysis of GWAS would have valuable implications. In addition, cross-ethnic GWAS has an advantage for fine-mapping because the linkage disequilibrium patterns that differ across populations can improve the resolution\(^5\). Thus, we report analyses of novel genome-wide association study data of 1173 sporadic ALS cases and 8925 normal controls in a Japanese population and meta-analysis with the largest ALS study in a European population\(^6\). We also validate the candidate region with a combined replication study using 707 other ALS cases and 971 controls from a Japanese population and a Chinese dataset\(^2\). Using a gene-based analysis, we identify four additional genes significantly associated with ALS. The discovery of novel risk genes advances our understanding of ALS aetiology.

## Results

### Genome-wide association analysis

To identify new susceptibility loci operating in sporadic ALS, we conducted a GWAS in a Japanese sample of 1173 sporadic ALS cases and 8925 controls (Supplementary Figs. 1, 2, and 3). The 56 SNPs passed the condition with \(p < 5.0 \times 10^{-6}\) (Supplementary Data 1), while no individual SNPs passed the genome-wide significant \(p\)-value threshold of \(5.0 \times 10^{-8}\). The data were then incorporated into a meta-analysis with a large-scale GWAS involving 20,806 patients diagnosed with ALS and 53,439 control subjects of European ancestry\(^6\). The Manhattan plot in Fig. 1 shows four identified loci in the European population that achieved genome-wide significance (Fig. 1a; loci include GPX3-TNIP1, C9orf72, TBK1, and UNC13A; Fig. 1b)\(^1,4\). In addition, we found the novel genome-wide statistical signature for SNPs in linkage disequilibrium on chromosome 10q25.2. This region spans several hundred kilobases and encompasses ACSL5 (Supplementary Data 2). Table 1 lists the top three SNPs displaying \(p < 5.0 \times 10^{-8}\) in the ACSL5 gene body region: rs58854276 (\(p = 2.97 \times 10^{-8}\); odds ratio (OR) = 1.080; 95%CI = 1.065–1.095), rs11195948 (\(p = 3.99 \times 10^{-8}\); OR = 1.079; 95%CI = 1.064–1.094), and rs3736947 (\(p = 3.61 \times 10^{-8}\); OR = 1.080; 95%CI = 1.065–1.095). The top three SNPs in ACSL5 are all intronic variants. Figure 2 shows the regional mapping of this region.

### Replication analysis

The association result was replicated in the combined meta-analysis with the Chinese dataset (1234 ALS cases and 2850 controls)\(^7\) and an independent new Japanese dataset (707 ALS cases and 971 controls). Among the former top three SNPs, two SNPs—rs11195948 and rs3736947—were validated (Table 1, Fig. 3, Supplementary Fig. 4). The \(p\)-value of rs11195948 was the most significant (\(p = 1.82 \times 10^{-14}\); OR = 1.136; 95%CI = 1.098–1.176). The tendency of the ratio in the Chinese sample was in the same direction (\(p = 2.09 \times 10^{-4}\); OR = 1.199; 95%CI = 1.103–1.290). The red horizontal line indicates the genome-wide significance threshold of \(p = 5 \times 10^{-8}\). The blue line indicates the suggestive threshold of \(p = 5 \times 10^{-7}\). The \(p\)-values of Japanese GWAS were calculated from the logistic regression model. For the combined dataset of the discovery and replication studies, the most significant SNP was rs3736947 (\(p = 7.81 \times 10^{-11}\); OR = 1.087; 95%CI = 1.073–1.101).

### Multi-ethnic meta-analysis

To maximise the power to detect genes related to ALS, we conducted the largest meta-analysis, with 23,213 cases and 71,579 controls from Japanese, European, and Chinese populations (Fig. 4a, b, and Supplementary Data 3). In the meta-analysis, the SNPs in ACSL5 also reached genome-wide significance: rs58854276 (\(p = 6.23 \times 10^{-10}\), rs11195948 (\(p = 6.23 \times 10^{-10}\), and rs3736947 (\(p = 3.99 \times 10^{-8}\), were validated (Fig. 4).
= 1.79 × 10⁻¹⁸), and rs3736947 (p = 1.84 × 10⁻¹¹). In addition, a novel locus in the long non-coding RNA TSBP1-AS1 reached genome-wide significance (rs140736091, p = 1.36 × 10⁻⁸), although this locus requires a replication study.

Functional analysis of ACSL5. A multi-tissue eQTL database GTEx v8 (https://www.gtexportal.org/home)⁸ revealed a significant relationship between rs3736947 and ACSL5 expression. The risk allele (C) is associated with higher ACSL5 expression than the non-risk allele (A) (p = 6.5 × 10⁻⁴⁹, normalised effect size = −0.39 in the whole-blood dataset). To confirm the association, we conducted gene expression analysis for ACSL5 by quantitative real-time PCR using lymphoblastoid B cell lines (LCLs) from age-matched and sex-matched Japanese patients with ALS, each genotype. The expression of ACSL5 was significantly higher in the LCLs of patients with the CC (n = 20) and AC (n = 20) genotype of rs3736947 than in those with the AA genotype (n = 20) (CC vs AA, p = 0.0035, AC vs AA, p = 0.0166, Kruskal-Wallis test followed by Steel-Dwass test, Fig. 5, Supplementary Data 4).

Gene-based association analysis. We performed gene-based association analysis in the largest meta-analysis set, with 23,213 cases and 71,579 controls from Japanese, European, and Chinese populations. The gene-based test confirmed that the already discovered genes, TNIP1 (p = 3.91 × 10⁻⁷), C9orf72 (p = 1.96 × 10⁻²¹), KIF5A (p = 1.06 × 10⁻⁷), and SCFD1 (p = 2.25 × 10⁻⁸)⁴. In addition, the novel genes ERGIC1 (p = 5.90 × 10⁻⁷), RAPGEF5 (p = 3.71 × 10⁻⁷), FNBP1 (p = 9.59 × 10⁻⁸), ACSL5 (p = 4.50 × 10⁻¹¹), and ATXN3 (p = 5.81 × 10⁻⁷), also reached genome-wide significance (Table 2, Fig. 6a, b, Supplementary Data 5).

Discussion
In this study, the region in ACSL5 was discovered as a novel risk locus for sporadic ALS by meta-analysis between Japanese and European datasets and was replicated in the Chinese dataset and another Japanese dataset. Expression analysis showed that the risk allele is associated with increased ACSL5 expression. The expression of ACSL5 mRNA in spinal motor neurons isolated by laser-capture microdissection in 12 sporadic ALS patients and nine controls was catalogued by Batra et al.¹⁰,¹¹. ACSL5 mRNA expression was possibly higher in sporadic ALS than in controls (Supplementary Fig. 5; p-value = 0.033 with Mann-Whitney U test). Similarly, another report showed that ACSL5 mRNA expression in the spinal anterior horn was upregulated in sporadic ALS patients compared with that in controls¹².

ACSL5 is one of the members of the acyl-CoA synthetase long chain family. Acyl-CoA synthetase produces acyl-CoA for numerous metabolic pathways, such as cellular lipid metabolism; transcriptional regulation; intracellular protein transportation; and protein acylation in various tissues, including skeletal muscle, the liver, and the brain.¹³ ACSL5 is a neurotoxic A1 astrocyte-related gene, and is up-regulated in A1 astrocytes.¹⁴ A1 astrocytes are abundant in various neurodegenerative diseases, including...
ALS, and they induce the death of neurons in the central nervous system. We speculated that increased expression of ACSL5 could induce A1 astrocytes, cause motor neuron death, and lead to ALS development. Another possible reason why ACSL5 is associated with ALS may be related to lipid metabolism. The risk allele (A) of rs58854276 has been reported to be associated with lower HDL-cholesterol in Japanese individuals. There have been several reports that dyslipidaemia increases the risk of ALS. However, the association between ALS and dyslipidaemia has not been replicated in other studies. Further studies are warranted to clarify the association between ACSL5 and ALS onset.

In the largest meta-analysis with 23,213 cases and 71,579 controls from Japanese, European, and Chinese populations, we identified a novel locus at 6p21, which reached genome-wide significance, in addition to ACSL5. The top SNP in the locus (rs140736091) was in the long non-coding RNA TSBP1-AS1. Some studies have reported that long non-coding RNAs are associated with ALS, but their role in ALS still needs to be elucidated. Therefore, further replication study and functional analysis will be needed to clarify the association between rs140736091 and patients with ALS.

The gene-based test for the largest multi-ethnic meta-GWAS from Japanese, Chinese and Europeans revealed novel genes, ERGIC1, RAPGEF5, FNB1, and ATXN3, in addition to ACSL5.

**Fig. 2 Regional ALS association plot of the ACSL5 locus from the meta-analysis results using LocusZoom.** The meta-analysis between European and Japanese GWAS shows rs58854276 as the SNP most strongly associated with ALS (p = 2.97 × 10^-8).

**Fig. 3 Forest plots showing the effects of rs3736947 in ACSL5 in each cohort and meta-analysis.** Forest plots showing the effects of rs3736947 on ALS in each cohort and meta-analysis.

**Fig. 4 Multi-ethnic meta-analysis among European, Japanese, and Chinese genome-wide association studies (GWAS).** a Manhattan plot of the multi-ethnic meta-analysis among European, Japanese, and Chinese GWAS. The novel loci ACSL5 and TSBP1-AS1 are shown in red. Loci identified in other studies are presented with lines along the gene names, i.e. GPX3-TNIP1, C9orf72, TBK1, and UNC13A. The red horizontal line indicates the genome-wide significance threshold of p = 5 × 10^-8. The blue line indicates the suggestive threshold of p = 5 × 10^-6. The genes indicated in red text colour with lines are novel genes and other genes are known risk genes. b Q–Q plot of the multi-ethnic meta-analysis among European, Japanese and Chinese GWAS.

ERGIC1 is a membrane-bound protein that is localised to the endoplasmic reticulum (ER)–Golgi intermediate compartment (ERGIC). The ERGIC mediates membrane traffic and selective transport of cargo between the ER and the Golgi complex. ER–Golgi transport dysfunction is reported to be a common pathogenic mechanism in SOD1-, TDP-43-, and FUS-associated ALS, and ER stress has been implicated in ALS aetiology. Combined GWAS of genetic overlap between ALS and fronto-temporal dementia-spectrum neurodegenerative diseases identified rs538622 near ERGIC1.

RAPGEF5 is a member of the Ras subfamily of GTPases, which function in signal transduction for cell growth and differentiation as guanosine diphosphate (GDP)/guanosine triphosphate (GTP)-regulated switches cycling between inactive GDP- and active GTP-bound states. RAPGEF5 has been reported to be associated with telencephalic neurogenesis. The RAPGEF5 transcript is expressed predominantly in the brain.
FNBP1, a member of the formin-binding protein family, is a membrane-associated protein. It plays an important role in clathrin-mediated endocytosis. FNBP1 is upregulated in the spinal cord of SOD1 G93A mice. ATXN3 is a ubiquitously expressed deubiquitinase that plays important roles in the ubiquitin proteasome system, transcriptional regulation and neuroprotection. A recent meta-analysis of European GWAS suggested an association between SNP (rs10143310) in ATXN3 and ALS, although the SNP did not achieve genome-wide significance (p = 3.2 × 10−7). The CAG repeat expansion in the coding region of ATXN3 causes spinocerebellar ataxia type 3 (SCA3). SCA3 patients and ALS patients have common pathologies, such as TDP-43-positive inclusions in the lower motor neurons of the anterior horn of the spinal cord and brainstem. A gene involved in cerebellar ataxia, ATXN2, has also been described as a risk gene for sporadic ALS. An intermediate repeat expansion in ATXN2 is associated with the risk of ALS.

In conclusion, multi-ethnic GWAS identified the association of the ACSL5 locus with ALS. This locus was identified by combining GWAS results from our Japanese dataset with the largest set of European GWAS data and was replicated in an independent Japanese cohort and a Chinese cohort. In addition, gene-based analysis identified ERGIC1, RAPGEF5, FNBP1, ACSL5, and ATXN3. While these genes reached the discovery stage of the analysis, further replication analysis or functional analysis in ALS is warranted. Nevertheless, the discovery of novel risk loci significantly advances our understanding of ALS aetiology.

Methods

Study subjects in the Japanese cohort. In the discovery cohort, we performed a GWAS in Japanese sporadic ALS cases from the Japanese Consortium for ALS research (JaCALS) and in normal controls from the Tohoku Medical Megabank Project (TMM). In the replication cohort, we obtained DNA from ALS patients registered by BioBank Japan and normal controls registered in the Pharma SNP Consortium. The ethics committees of the respective research projects approved this study. Written informed consent for this study was obtained from all the participants.

DNA extraction from ALS patients. To extract genomic DNA, peripheral whole-blood samples were processed using an Autopure LS system (Qiagen), for automated nucleotide purification according to the manufacturer’s instructions. We omitted RNase treatment, measured the concentration of the double-stranded DNA with PicoGreen (Life Technologies, Carlsbad, CA, USA), and adjusted the concentration of the DNA to 200 ng/µl in Elution Buffer (Qiagen).

Table 2 Summary of significant genes in the multi-ethnic meta-analysis of gene-based association analysis among Japanese (JaCALS), European, and Chinese cohorts.

| Ensembl Gene ID     | Chr | Start   | End     | No of SNPS | NPARAM | N  | ZSTAT     | P       | SYMBOL | Reported        |
|---------------------|-----|---------|---------|------------|---------|----|-----------|---------|--------|-----------------|
| ENSG00000145901     | 5   | 150409506 | 150473138 | 126        | 23      | 94792 | 4.9399    | 3.91 × 10⁻⁷ | TNIP1   | Benjamini, B. et al. (2017) (as GPX2-TNIP1)⁷ |
| ENSG0000013719      | 5   | 172261278 | 172379688 | 140        | 31      | 94792 | 4.859     | 5.90 × 10⁻⁷ | ERGIC1  | Novel           |
| ENSG00000136237     | 7   | 22157856  | 22396763  | 385        | 65      | 94792 | 4.9503    | 3.71 × 10⁻⁷ | RAPGEF5 | Novel           |
| ENSG00000120162     | 9   | 27325207  | 27529779  | 304        | 49      | 94792 | 9.4345    | 1.96 × 10⁻¹ | MOB8B   | van Es, M. A. et al. (2009) (as C9orf72)⁸ |
| ENSG00000147896     | 9   | 27524312  | 27526496  | 5          | 2       | 94792 | 7.3632    | 8.98 × 10⁻¹ | IFN1    | van Es, M. A. et al. (2009) (as C9orf72)⁸ |
| ENSG00000147894     | 9   | 27546544  | 27573684  | 59         | 9       | 94792 | 7.399     | 6.86 × 10⁻⁴ | IFN1    | van Es, M. A. et al. (2009) (as C9orf72)⁸ |
| ENSG00000187239     | 7   | 132649466 | 132805834 | 100        | 13      | 94792 | 5.6192    | 9.59 × 10⁻⁹ | FNBPI   | Novel           |
| ENSG00000197142     | 10  | 114133776 | 11418138  | 49         | 7       | 94792 | 6.4829    | 4.50 × 10⁻¹ | ACSL5   | as ACSL5         |
| ENSG00000023041     | 10  | 114900034 | 114906762 | 12         | 2       | 94792 | 5.1368    | 1.40 × 10⁻⁶ | ZDHHC6  | Nicolas, A. et al. (2018) (as KIF5A)⁹ |
| ENSG00000035454     | 12  | 58017193  | 58027138  | 13         | 3       | 94792 | 5.1883    | 1.06 × 10⁻⁵ | B4GALNT1 | Novel           |
| ENSG00000092134     | 4   | 31028329  | 31089269  | 55         | 8       | 94792 | 4.944     | 3.83 × 10⁻⁷ | G2E3    | van Rheenen, W. et al. (2016) (as SCFD1)¹⁰ |
| ENSG00000092108     | 4   | 31093138  | 31065018  | 137        | 10      | 94792 | 4.5869    | 2.25 × 10⁻⁶ | SCFD1   | van Rheenen, W. et al. (2016)¹⁰ |
| ENSG00000066427     | 14  | 92524896  | 92572695  | 98         | 5       | 94792 | 4.8619    | 5.81 × 10⁻⁷ | ATXN3   | Novel           |

Chr: chromosome, Start: start position of gene, End: end position of gene. Position is based on Human Genome Assembly build 37. No of SNPs: The number of SNPs annotated to that gene that were found in the data and were not excluded based on internal SNP QC. NPARAM: the number of relevant parameters used in the model. N: the sample size used when analysing that gene. P: the gene p-value, using asymptotic sampling distribution. SYMBOL: gene name. Significant threshold of p = 2.85 × 10⁻⁶ (→0.05/17544).
The blue line indicates the suggestive threshold of among European, Japanese, and Chinese cohorts.

Fig. 6 The gene-based analysis for the multi-ethnic meta-analysis among European, Japanese, and Chinese cohorts. a Manhattan plot of the gene-based analysis for the multi-ethnic meta-analysis among European, Japanese, and Chinese cohorts. The red horizontal line indicates the genome-wide significance threshold of \( p = 5 \times 10^{-8} \) (\( \chi^2 = 7.85 \times 10^{-8} \)). The blue line indicates the suggestive threshold of \( p = 2.85 \times 10^{-6} \). The genes indicated in red are novel genes, and the other genes are known risk genes. b Q-Q plot of gene-based analysis for the multi-ethnic meta-analysis among European, Japanese, and Chinese cohorts.

Japanese ALS patients for the discovery study. The case cohort comprised 1,245 candidate sporadic ALS patients recruited from the Japanese Consortium for ALS research (JaCALS), which included 32 neurology facilities in Japan. Patients with a family history for ALS were excluded. The included patients were diagnosed with definite, probable, probable laboratory-supported, or possible ALS based on the revised El Escorial criteria\(^37\) and were of Japanese ancestry. The DNA samples used in this study were collected from February 2006 to December 2017. The approval of the overall research was obtained from the Ethics Committee of Nagoya University with number 2004-0181-2. The ethics committees of all participating institutions approved the study. Details of the JaCALS have been described elsewhere\(^38\). All 1,245 sporadic ALS patients were genotyped with Illumina (San Diego, CA, USA) HumanOmniExpressExome-8 BeadChips (version 1.0, 1.2, and 1.3).

Japanese controls for the discovery study. The control cohort was the prospective cohort of the Tohoku Medical Megabank Project (TMM)\(^39,40\). From these TMM samples, we used 9,924 samples genotyped with Illumina HumanOmniExpressExome-8 BeadChips (version 1.2-B) at RIKEN. Approval to use the dataset as the control was obtained from the Ethical Committee of the Tohoku University Graduate School of Medicine, the TMM, and the Materials and Information Distribution Review Committee of Tohoku Medical Megabank Organization.

Quality control of genotyped SNPs and imputation. For the combined genotyped dataset with 1,245 cases and 9,924 controls, we selected SNPs with minor allele frequency \( \geq 0.03 \), Hardy–Weinberg equilibrium \( \geq 0.001 \), and genotype call rate \( \geq 95\% \). This step retained 516,447 SNPs from 936,632 SNPs. The qualified genotype dataset was imputed with the 2,037 whole-genome reference panel (2KJPN) from TMM\(^38\) using Impute 4\(^39\). The target individuals themselves for the imputation were not included in the 2KJPN reference panel\(^40\). According to the protocol by Shiodo et al.\(^41\), the imputed genotype data in Oxford GEN format were converted to Plink BED format by selecting the genotype with the highest posterior probability for each SNP and individual. In the conversion, highest posterior probabilities less than 0.9 were handled as missing genotypes. Finally, we constructed the imputed dataset with 1245 cases and 9924 controls from 27,893,192 SNPs.

Pre-processing and GWAS. We excluded 69 suspected ALS patients and two non-ALS patients from 1,245 cases. Samples that were identical-by-descent (\( \leq 0.1875 \)) among 1174 cases and 9924 controls were removed. The thresholds were based on other published studies\(^42-45\). Finally, 1173 cases and 8925 controls were used for GWAS. For GWAS, we selected 4,349,201 SNPs with minor allele frequency \( \geq 0.03 \), Hardy–Weinberg equilibrium \( \geq 0.001 \), and genotype call rate \( \geq 95\% \). We conducted logistic regression analysis with principal components 1–20 from the principal component analysis as covariates using Plink (version 1.90b5.1).

Meta-analysis of Japanese and European dataset. We downloaded the summary statistics of the 20,806 ALS cases and 59,804 controls in the European dataset from the website https://als.unimedsb.br/ims. Inverse-variance meta-analysis was conducted between the 20,806 ALS cases and 59,804 controls in the European dataset\(^6\) and the 1173 cases and 8925 controls of Japanese ancestry using METAL (version 2011-03-25)\(^46\). The regions from 114,000,000 to 114,300,000 in chromosome 10, including the significant SNPs in ACSL5, were plotted using LocusZoom (version 1.4) with --build hg19 --p Glasgow --source 1000G_Nov2014 options\(^47\).

Japanese replication cohort. We obtained DNA from 707 ALS patients registered by BioBank Japan\(^35,36\) and 971 normal controls registered in the Pharma SNP Consortium. Twenty-nine SNPs with a suggestive threshold of \( p = 5 \times 10^{-6} \) on chromosome 1q25-2 were genotyped via the multiplex PCR-based target sequencing method. Primers were designed using Primer 3 software. The products of multiplex PCR were sequenced using Illumina HiSeq 2500 (Illumina Inc., San Diego, CA, USA). The sequence data were analysed using a standard pipeline. The details of the methods have been previously described\(^48\).

For this analysis, we did not have genome-wide genotype information for these samples. We conducted logistic regression analysis using Plink with no covariate, e.g. principal components, for these 29 SNPs.

Meta-analysis of Chinese and Japanese replication cohort. We downloaded the summary statistics of the 1234 ALS cases and 2850 controls in the Chinese cohort\(^7\). Inverse-variance meta-analysis was conducted for the 1234 ALS cases and 2850 controls in the Chinese cohort and 707 ALS cases and 971 normal controls in the Japanese replication cohort using METAL (version 2011-03-25)\(^46\). A value of \( p < 5 \times 10^{-4} \) (\( \chi^2 = 0.05/29 \); Bonferroni adjusted) in the replication stage was considered statistically significant.

Multi-ethnic meta-analysis. Inverse-variance meta-analysis was conducted among 1234 cases and 2850 controls in the Chinese cohort\(^7\) and the former samples in European\(^6\) and Japanese datasets (JaCALS) using METAL (version 2011-03-25)\(^46\).

Gene-analysis analysis of the multi-ethnic dataset. For the summary statistics data of the multi-ethnic meta-analysis of Japanese, European, and Chinese datasets with 3,704,464 SNPs, a gene-based analysis was applied using MAGMA (version 1.07) with the default option\(^49\). In the pre-processing step, each SNP was checked to be mapped to a specific gene. If the SNP was located within the gene body region of a gene, the SNP was annotated to this gene, i.e. a SNP in an intergenic region was not annotated to any gene. In the pre-processing step, an annotation window range can be set to include the peripheral regions around the genes. The default value of the annotation window range is zero in MAGMA (version 1.07). The value is the strictest option in this pre-processing step since extra regions around the gene are not included in gene-based analyses. Thus, we used the default option. After the pre-processing, 1,585,558 SNPs in total were annotated to 17,544 genes. For the gene-based analysis, a \( p \)-value should be calculated for each gene from the SNPs annotated to this gene in the pre-processing step. We used the default SNP-wise mean model in MAGMA for the calculation step. The SNP-wise mean model (the mean of the \( \chi^2 \) statistic for the SNPs in a gene) is highly similar to the commonly used SKAT model (with inverse variance weights)\(^50\). The drawback of this method is that it decreases the power to detect associations for rare variants. In our analysis, the minor allele frequency in the discovery Japanese cohort was more than 0.03, and the problem was negligible.
Finally, 13 genes with a genome-wide significance threshold of $p = 2.85 \times 10^{-6}$ ($\alpha/0.05=17544$, after Bonferroni correction) were selected.

### Quantitative real-time PCR

Lymphoblastoid B cell lines (LCLs) were prepared from peripheral blood B cells of ALS patients using standard Epstein-Barr virus transformation techniques at the time of registration for JaCALS. LCLs obtained from 20 patients with age-matched and sex-matched ALS (Supplementary Table 1) for each genotype of rs3736947 were applied to the gene expression analysis. All 60 patients whom we selected to quantify eQTL were diagnosed with sporadic ALS. Further, we performed exome sequencing or target resequencing for all 60 patients who were selected to quantify eQTL. The details of the methods have been previously described. These patients had no pathogenic variant in the ALS-related genes, such as SOD1, FUS, and TARDDBP. Total RNA was extracted from the LCLs using the PureLink RNA Mini kit (Thermo Fisher Scientific, Waltham, MA, USA). Total RNA was transcribed using the SuperScript IV VILO Master Mix (Thermo Fisher Scientific). Real-time PCR was performed using the THUNDERBIRD SYBR qPCR Mix (Toyobo, Osaka, Japan) and the CFX96 system (BioRad, Hercules, CA, USA), according to the manufacturer’s instructions. The expression level of the internal control, β2 microglobulin, was simultaneously quantified. The primers are listed in Supplementary Table 2. Differences in AC5.5 mRNA expression among three genotype groups were evaluated by the Kruskal–Wallis test followed by the Steel–Dwass test. Statistical analyses were conducted using the JMP15.1 program (SAS Institute Inc., Cary, NC, USA).

### Data availability

The summary statistics of our genome-wide association studies are available at the Human Genetic Variation Database (Accession ID: HGV0000013). The source data of all relevant data are available from G.S. upon request.

### References

1. van Es, M. A. et al. Amyotrophic lateral sclerosis. Lancet 390, 2084–2099 (2017).

2. Nakamura, R. et al. Next-generation sequencing of 28 ALS-related genes in a Japanese ALS cohort. Neurobiol. Aging 39, 219 e211–218 (2016).

3. Al-Chalabi, A. et al. An estimate of amyotrophic lateral sclerosis heritability using twin data. J. Neurol. Neurosurg. Psychiatry 81, 1324–1326 (2010).

4. van Rheenen, W. et al. Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat. Genet. 48, 1043–1048 (2016).

5. Peterson, R. E. et al. Genome-wide association studies in ancestrally diverse populations: Opportunities, methods, pitfalls, and recommendations. Cell 179, 589–603 (2019).

6. Nicolas, A. et al. Genome-wide analyses identify KIF5A as a novel ALS gene. Neuron 97, 1268–1283 e1266 (2018).

7. Benyamin, B. et al. Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. Nat. Commun. 8, 611 (2017).

8. GTEx Consortium. The genotype-tissue expression (GTEx) project. Nat. Genet. 45, 580–585 (2013).

9. van Es, M. A. et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p22.12 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat. Genet. 41, 1083–1087 (2009).
Acknowledgements

The authors thank all of the ALS patients who participated in this study. The authors also thank all of the participating doctors and staff of the Japanese Consortium for Amyotrophic Lateral Sclerosis research (JaCALS). The participating doctors of JaCALS are thanked all of the participating doctors and staff of the Japanese Consortium for Amyotrophic Lateral Sclerosis research (JaCALS). The participating doctors of JaCALS are

Author contributions

R.N., Ka.M., G.T., M.Nak., M.Nag., and G.S. conceived the study and interpreted the results. R.N., Ka.M., G.T., M.Nak., S.F., and M.Nag. performed statistical and bioinformatic analyses. R.N., N.A., N.A., D.Y., H.A., H.W., H.L., Ma.K., Y.I., K.K., No.H., M.M., A.T., O.K., M.O., K.S., K.S., N.S., M.A., Ya.O., T.Y., K.A., R.H., I.A., K.O., Ko.M., K.H., T.I., O.O., K.N., K.K., and G.S. were involved in sample collection and management of Japanese ALS cohort (JaCALS). Ka.M., S.F., N.I., N.M., and M.Nag provided Japanese GWAS control sample for the discovery cohort. Y.K., S.I., Y.M., and Mi.K. provided Japanese ALS sample and control sample for the replication cohort. Y.K., S.I., Y.M., and Mi.K. performed genotyping. G.T. and Ye.O. performed mRNA expression analysis. R.N., Ka.M., G.T., M.Nak., N.A., M.Nag., and G.S. wrote the first draft of the manuscript. All authors contributed to the revision of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s42003-020-01251-2.

Correspondence and requests for materials should be addressed to M.Nag or G.S.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020
