A large genome scan for rare CNVs in amyotrophic lateral sclerosis

Hylke M. Blauw,1 Ammar Al-Chalabi,2 Peter M. Andersen,4,5 Paul W.J. van Vught,1 Frank P. Diekstra,1 Michael A. van Es,1 Christiaan G.J. Saris,1 Ewout J.N. Groen,1 Wouter van Rheenen,1 Max Koppers,1 Ruben van ‘t Slot,6 Eric Strengman,6 Karol Estrada,7,8 Fernando Rivadeneira,7,8 Albert Hofman,6 Andre G. Uitterlinden,7,8 Lambertus A. Kiemeneij,9 Sita H.M. Vermeulen,9 Anna Birve,4 Stefan Waibel,5 Thomas Meyer,10 Simon Cronin,11 Russell L. McLaughlin,11 Orla Hardiman,11,12 Peter C. Sapp,11,14,15,16 Martin D. Tobin,17 Louise V. Wain,17 Barbara Tomik,18 Agnieszka Slowik,18 Robin Lemmens,19,20,21 Dan Rujescu,22 Claudia Schulte,23 Thomas Gasser,23 Robert H. Brown Jr.,13,14 John E. Landers,13,14 Wim Robberecht,20,21,22 Albert C. Ludolph,5 Roel A. Ophoff,2,24 Jan H. Veldink,1,25 and Leonard H. van den Berg1,25,*

1Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands
2Department of Medical Genetics and Rudolf Magnus Institute, Utrecht 3584 CX, The Netherlands
3Department of Clinical Neuroscience, King’s College London, Medical Research Council (MRC) Centre for Neurodegeneration Research, Institute of Psychiatry, London SE5 8AF, UK
4Department of Clinical Neuroscience, Umeå University, Umeå SE-901 85, Sweden
5Department of Neurology, University of Ulm, Ulm 89081, Germany
6Complex Genetics Section, Division of Biomedical Genetics, Department of Medical Genetics, University Medical Centre Utrecht, 3584 CX Utrecht, The Netherlands
7Department of Internal Medicine, Erasmus University Medical Centre Rotterdam, Rotterdam 3015 GE, The Netherlands
8Department of Epidemiology Erasmus University Medical Centre Rotterdam, Rotterdam 3015 GE, The Netherlands
9Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen 6500 HB, The Netherlands
10Department of Neurology, Charité University Hospital, Humboldt-University, Berlin 13353, Germany
11Department of Neurology, Beaumont Hospital, Dublin, Ireland
12Trinity College Institute of Neurosciences, Trinity College, Dublin, Ireland
13Department of Neurology, Massachusetts General Hospital, Charlestown, MA 02129, USA
14Department of Neurology, University of Massachusetts School of Medicine, Worcester, MA 01655, USA
15Howard Hughes Medical Institute, Cambridge, MA, 01605, USA
16Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 01605, USA
17Department of Health Sciences and Department of Genetics, University of Leicester, Leicester LE1 7RH, UK
18Department of Neurology, Jagiellonian University, Krakow 31-007, Poland
19Experimental Neurology, University of Leuven, Leuven 3000, Belgium
20Vesalius Research Centre, Flanders Institute for Biotechnology (VIB), Leuven 3000, Belgium
21Department of Neurology, University Hospital Leuven, University of Leuven, Leuven 3000, Belgium
22Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich 80336, Germany
23Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and German Center for Neurodegenerative Diseases, Tübingen 72074, Germany
24UCLA Center for Neurobehavioral Genetics, Los Angeles 60095, USA
25These authors contributed equally to this work
* To whom correspondence should be addressed: Department of Neurology, UMC Utrecht, Heidelberglaan 100, 3584 CX, Utrecht, The Netherlands; Telephone: 0031 88 7555555; Fax: 0031 30 2542100; E-mail: l.h.vandenberg@umcutrecht.nl;
ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease selectively affecting motor neurons in brain and spinal cord. Recent genome-wide association studies have identified several common variants which increase disease susceptibility. In contrast, rare copy-number variants (CNVs), which have been associated with several neuropsychiatric traits, have not been studied for ALS in well-powered study populations. To examine the role of rare CNVs in ALS susceptibility, we conducted a CNV association study including over 19,000 individuals. In a genome-wide screen of 1,875 cases and 8,731 controls, we did not find evidence for a difference in global CNV burden between cases and controls. In our association analyses we identified two loci that met our criteria for follow-up: the \textit{DPP6} locus (OR=3.59, \textit{p}=6.6x10^{-3}), which has already been implicated in ALS pathogenesis, and the 15q11.2 locus, containing \textit{NIPA1} (OR=12.46, \textit{p}=9.3x10^{-5}), the gene causing hereditary spastic paraparesis (HSP) type 6. We tested these loci in a replication cohort of 2,559 cases and 5,887 controls. Again, results were suggestive of association, but did not meet our criteria for independent replication: \textit{DPP6} locus: OR=1.92, \textit{p}=0.097, pooled results: OR=2.64, \textit{p}=1.4x10^{-3}; \textit{NIPA1}: OR=3.23, \textit{p}=0.041, pooled results: OR=6.20, \textit{p}=2.2x10^{-5}). Our results highlight \textit{DPP6} and \textit{NIPA1} as candidates for more in-depth studies. Unlike other complex neurological and psychiatric traits, rare CNVs with high effect size do not play a major role in ALS pathogenesis.
INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterised by the selective death of motor neurons in the spinal cord, brainstem and cortex, leading to progressive paralysis and eventually death. The peak incidence lies between age 50 and 75 and the average time of survival is about 3 years after onset of symptoms (1). There is no curative therapy. Approximately 5-10% of patients have a family history of ALS (FALS), of which <25% are explained by known genes (2). In the majority of ALS patients, however, there is no family history and the disease is said to be sporadic. Sporadic ALS is considered to be a complex trait, in which multiple genetic and environmental risk factors increase disease susceptibility. Concordance rates from twin studies suggest that the heritability is between 0.38 and 0.85 (3). Despite this relatively high heritability, disappointingly few well-established genetic risk factors have been discovered.

Recently, genome-wide association studies (GWAS) have identified a small number of loci that increase disease susceptibility, including *UNC13A* and the 9p21.2 locus (4). These studies are designed to detect association of common genetic variants with usually relatively mild effects. It is likely, however, that rare alleles will explain an important part of the heritability in most complex traits (5). Rare copy-number variants (CNVs) have been associated with several neuropsychiatric traits, such as autism, schizophrenia and epilepsy (6-8). We previously reported on CNVs in ALS patients, but these studies were not powered to detect association of rare events (9-11).

In order to systematically search for CNVs that confer susceptibility to ALS, we performed an international collaborative study focusing on rare CNVs. We conducted a genome-wide screen for CNVs in 1,875 cases and 8,731 controls and sought to replicate identified loci in a second cohort of 2,559 cases and 5,887 controls. Overall, we included over 19,000 individuals.
RESULTS

After applied quality control filters we included 1,875 ALS cases and 8,731 controls in the genome-wide screen (Table 1 and Table S1). We first examined global CNV burden and differences in genic content of CNVs with Plink software (12). Characteristics of detected CNVs are shown in Table S2. In contrast to previous reports (9, 13) we did not find evidence of a higher global CNV burden in ALS cases than in controls or a difference in genic content in ALS cases for CNVs in general, or for large (>500kb) CNVs (Table S2). We found 292 genes that were affected by CNVs in ALS cases and not in controls in this study or in the Database of Genomic Variants (excluding BAC array-based studies). These were present in 169 CNVs, belonging to 159 unique individuals. These might be either ALS-specific pathogenic variants or may represent extremely rare benign variants without any association with disease. These variants do not include known ALS candidate genes. As a reference for future studies we have included these genes in Table S3.

To identify associated loci we used a gene-based approach focusing on rare CNVs (Material and Methods). Each gene affected by one or more CNVs was tested for association with deletions and duplications lumped and with deletions and duplications separately. 4,991 genes were affected by one or more CNVs (i.e. deletions or duplications). Of these genes, 2,609 were affected by one or more deletions and 3,259 genes by duplications. The total number of tested hypotheses is therefore 10,859. However, because most CNVs are very rare, the majority of the tested hypotheses will not produce significant results a priori. For example, only 48 regions contained CNVs that occurred in a total frequency (cases and controls combined) of more than 0.25% (not shown).

We identified 3 genomic regions containing genes with p-values <0.01 and a frequency in controls of <1%: the DPP6 locus at 7q36.2, and the 15q11.2 and 19q13.2 loci, which both contain multiple genes (Table S4). Only one locus was overrepresented in controls using the same criteria (8q24.3 locus, p=2.5x10^{-6}) (not shown). For replication purposes, we focused on the three regions that were overrepresented in ALS cases. We analysed available DNA samples with TaqMan qPCR to validate 28 CNVs in these three loci (Material and Methods). The three CNVs in the 19q13.2 locus, consisting of 3 putative deletions in ALS cases and none in controls, could not be validated and were thus regarded as false positives. All CNVs in the DPP6 and 15q11.2 loci (25 out of 25) were experimentally validated. The results for the two validated regions are shown in Table 2 and the
genomic organization is depicted in Figure 1. A breakdown per country and per platform of CNVs found in the two regions of interest can be found in Table S7.

In the DPP6 locus, the signal was driven mainly by duplications in the 5’ end of the gene (Figure 1). This region is a known CNV region, flanked by segmental duplications, which are known to mediate genomic rearrangements (14). Results for the DPP6 locus do not appear to be population-specific (Table S5). This is suggested by a non-significant p value for Woolf’s test (for heterogeneity of the odds ratios) (p=0.60) and reflected by the results after stratification by country of origin (p=6.6x10^{-3}, Cochran-Mantel-Haenszel test, Table 2).

The second locus is at 15q11.2 in the Prader-Willi/Angelman region. Duplications in this region did not contribute to the association, while deletions showed a strong signal (lowest P 9.3 x 10^{-5}, Fisher exact test, OR = 12.46) (Tables 2 and S6). This region is also a known CNV region flanked by segmental duplications (Figure 1). The signal in this region is caused by deletions with different breakpoints, overlapping four adjacent genes. The strongest association was, however, found for NIPA1 (Non-imprinted in Prader-Willi/Angelman syndrome region protein 1). The stratified analyses for the 15q11.2 locus showed more conservative results compared to the unadjusted Fisher's exact test, but there was no significant heterogeneity of the odds ratios across countries (p=0.98, Woolf’s test) (Table S6).

We examined the study populations for possible ancestry mismatch between cases and controls within countries, which could cause spurious associations. In the Dutch we identified 35 outliers in the MDS cluster plot (Figure S2), which were removed and the association analysis were repeated. Removal of the 35 samples produced similar p-values for NIPA1 (p= 9.4 x 10^{-5}, Fisher exact test, p=2.6x10^{-3}, Cochran-Mantel-Haenszel test) and slightly more modest for DPP6 (p= 9.1 x 10^{-3}, Fisher exact test, p=1.2x10^{-2}, Cochran-Mantel-Haenszel test) (not shown). Using the IBS-test as implemented in PLINK we found evidence for stratification in the Swedish (p=1.0x10^{-5}), but not in the Dutch, Belgians and Irish (Table S8 and Figure S2). The removal of the Swedish population from the analysis causes reduced statistical power, but again showed evidence for association for the two loci: NIPA1 (p= 2.6 x 10^{-3}, Fisher exact test, p=0.016, Cochran-Mantel-Haenszel test); DPP6 (p= 0.011, Fisher exact test, p=0.014, Cochran-Mantel-Haenszel test) (not shown).

Previously, we and others have reported on the association of a SNP (rs10260404) in DPP6 with ALS susceptibility. We considered whether the associated SNP serves as a proxy for the CNVs in the 5’ end of
DPP6, such that the results obtained in this study essentially reflect the same underlying signal as in previous studies. However, this is unlikely: rs10260404 is located in intron 3 of DPP6, which is more than 500 kb downstream of the CNV locus and the two loci are separated by several recombination hotspots. We found that there is no evident LD between rs10260404 and the CNVs: $D' = 0.16$, $R^2 = 0$ (Figure S3). The signal from the CNV locus is thus not explained by rs10260404 tagging the CNVs and represents an independent finding.

To further examine the significance of the findings in our genome-wide screen, we tested the two loci in a second cohort of 2,559 ALS cases and 5,887 controls (Tables 1 and S1). Results from our replication cohort again showed more CNVs in ALS cases compared to controls in both loci, but with smaller effect sizes than estimated in the genome-wide phase (Table 2): CNVs affecting DPP6 were found in 10 of 2,559 ALS cases and in 12 of 5,887 controls (0.39% vs. 0.20%, OR=1.92, p=0.097, Fisher exact test). Stratified analysis to correct for country showed a higher p value: p=0.24, Cochran-Mantel-Haenszel test). Deletions in NIPA1 were found in 7 of 2,559 cases and in 5 of 5,887 controls (0.27% vs. 0.08%, OR=3.23, p=0.041, Fisher exact test). Stratification per country showed similar (but more conservative) results for NIPA1 (p=0.055, one-sided Cochran-Mantel-Haenszel test). A breakdown of results per country can be found in tables S5 and S6. For the replication phase, we considered p<0.025 as statistically significant (Bonferroni correction of $\alpha=0.05$ for 2 loci), and so we do not consider the results from our initial genome-wide screen to have been replicated.

Given the estimated effect size in the initial genome-screen, our replication cohort was well-powered to detect associations for both loci at $\alpha=0.025$: 77% for the DPP6 locus and 93% for NIPA1. With the effect size from the replication study, however, power is reduced substantially: power was 30% to detect an association for DPP6 and 46% for NIPA1 (Figure S4). Therefore, while still compatible with association of both loci, our replication data indicate a more modest effect than initially estimated in the genome-wide discovery phase.

Recently, deletions in the 15q11.2 locus have been associated with schizophrenia and idiopathic generalised epilepsy (7, 8). We assessed whether ALS patients in our study population suffered from any additional co-morbidity (Table S9, Supplementary Material). With the exception of one patient with transient epileptic episodes, we did not find evidence of psychiatric or other neurologic co-morbidity in patients carrying a deletion in the 15q11.2 locus. Also, patients generally exhibited a classic ALS phenotype without a particularly long or short survival.
DISCUSSION

We have conducted a large-scale CNV association study including over 19,000 individuals to identify loci associated with ALS. Our results highlight two potential susceptibility genes for ALS. DPP6 was already considered a candidate gene for ALS, while NIPA1 has not previously been linked with ALS pathogenesis.

Previously, we reported the association of a SNP in DPP6 with ALS susceptibility, which has subsequently been replicated in other studies (15-18) but not corroborated by all (19-22). Thus, the role of DPP6 in ALS pathogenesis remains unclear. The SNP that was reported in the prior studies is not in LD with the CNVs we identified; therefore, our results independently point to DPP6 as a candidate gene for ALS. DPP6 (encoding dipeptidyl-peptidase 6) modulates the function and expression of potassium channels and excitability at the glutamatergic synapse (23) and is predominantly expressed in the central nervous system. DPP6 has also been identified as a potential candidate for autism (6). The signal we found in the present study is mainly driven by duplications in the 5’ end of the gene. The functional impact of these duplications is not known. They may alter or truncate transcripts or may disrupt regulatory elements in this region that alter gene expression (24). How this would increase susceptibility to ALS is not known and should be the subject of further studies.

The 15q11.2 locus contains four genes, of which NIPA1 showed the strongest signal and is the most plausible gene to explain an association from a biological point of view: mutations in NIPA1 cause hereditary spastic paraplegia type 6 (HSP 6), a neurodegenerative disease characterized by the selective degeneration of (upper) motor neurons (25). The same mutations cause progressive paralysis in C. elegans (26). Recently, CNVs in the same locus have been associated with schizophrenia and idiopathic generalized epilepsy (7, 8). These pleiotropic effects might be conferred by the different genes affected by the same CNVs, together with additional genetic variants at other loci. Notably, patients in our study with a deletion in the 15q11.2 locus did not have idiopathic generalized epilepsy, schizophrenia or any neurodevelopmental features, indicating that additional risk factors are required to cause a disease phenotype.

If NIPA1 is indeed involved in ALS pathogenesis, it may point to a common pathway that is involved in pathogenesis of both ALS and HSP. Although these are clinically different entities, they share important clinical features caused by selective motor neuron degeneration. Intriguing in this respect is the recent finding that mutations in SPATACSIN, causing HSP type 11, also cause autosomal recessive juvenile amyotrophic lateral sclerosis (27). Additionally, anecdotal reports suggest that SPAST mutations, causing HSP type 4, can also cause...
an ALS phenotype (28, 29). Both NIPA1 and SPAST are inhibitors of BMP (bone morphogenic protein) signalling (30). Therefore, this pathway may be important for motor neuron biology, and other genes in this signalling cascade may be targets for systematic assessment, such as resequencing studies to screen for rare variants.

In our study, we focused on gene-containing CNVs as a necessary means of prioritization of loci for follow-up. This approach will consequently not include CNVs without (known) genes, but which could still have major functional consequences. For instance, micro RNAs are likely to be involved as ALS modifiers (31), but these have not been assessed in this study and will have to be examined in more detail in the future. Also, most small (e.g. <50 kb) CNVs will probably go undetected in our study, because of limitations to the resolution of the used platforms.

We found evidence for population stratification in the Swedish population of the discovery phase, which should be kept in mind when interpreting the results. Ancestral mismatch can cause spurious associations, because of differences in allele frequencies of populations that are derived from different ancestral pools. We cannot rule out the possibility that the stratification in the Swedish population contributes to the overall association in the discovery phase. However, the removal of the Swedish from the analyses (with an inherent reduced statistical power) still reproduced the associations, which suggests that the results are not driven by stratification of the Swedish population.

The lack of true replication of the findings in the discovery phase can in our view be attributed to either type I error in the discovery phase or an overestimation of the effect size, and therefore a lack of statistical power in the replication phase to detect the association. The replication results illustrate that in the case of a true association of the 2 loci, the estimated effect from the initial genome-wide phase is probably an over-estimate of the true effect. Consequently, the replication cohort may not be adequately powered to detect the true (more modest) effect. Inflation of effect size in discovery phases of genome-screens is not uncommon and might explain a part of non-replications in genome-wide studies, including in ALS (32). Concluding from GWAS studies in ALS from the last several years, in our opinion the strongest evidence for implication in ALS pathogenesis exists for the UNC13A and 9p21.2 loci, which both have been convincingly replicated after identification in a genome screen (19). For the DPP6 locus, as discussed previously, the true role in ALS remains unclear, with conflicting results in different replication studies. For the remaining variants that have
emerged as candidates from GWAS, the evidence for true association remains low, most probably largely attributable to underpowered discovery cohorts (33-36), emphasizing the need for well-powered, internationally conducted studies. CNV studies in other complex disorders, like schizophrenia and epilepsy (7, 8), have identified CNV associations that confer relatively high effect sizes (odds ratio between ~5-25). Our study was adequately powered to detect these sorts of associations, but would not have detected variants with a more moderate effect. The growing body of ALS and control genotyping data will probably enable the identification of these variants in even larger samples in the future.

In conclusion, we have carried out a large CNV association study in ALS, but could not detect CNVs with high effect size such as found in other neurological and psychiatric complex traits. Our results highlight two loci with moderate effect size that may be implicated in ALS pathogenesis, and justify more in-depth examination, as they might point towards novel pathways that may be therapeutic targets for treating this devastating disease.

MATERIAL AND METHODS

Study populations

In this study we used genome-wide data and DNA samples from ALS cases and unaffected controls from The Netherlands, Belgium, Sweden, Denmark, Norway, Finland, Ireland, Germany, USA and Poland (Table 1 and Table S1). All ALS patients were diagnosed by neurologists specialized in neuromuscular diseases, and fulfilled the 1994 El Escorial criteria for probable or definite ALS (37). Patients with a family history of ALS were excluded from this study. The relevant medical ethical committees approved all studies and all participants provided informed consent.

ALS cases from The Netherlands were recruited from outpatient clinics at the national referral centers for ALS in The Netherlands: the University Medical Center Utrecht, the Radboud University Nijmegen Medical Center and the Academic Medical Center Amsterdam. Controls were derived from different sources: 1) healthy volunteers accompanying patients were recruited in the Neurology outpatient clinic in the University Medical Center Utrecht (9, 15). 2) 578 controls were recruited as part of a nationwide prospective study on motor neuron diseases in The Netherlands (4). 3) 7,732 controls were included from the three cohorts of the Rotterdam Study, a prospective population-based cohort study (38). 4) 1,622 controls were included from a genome-wide
association study on urinary bladder cancer in the Radboud University Nijmegen Medical Center (39). 5) 430 controls were included from a Dutch genome-wide association study on schizophrenia (7, 40) Belgian ALS cases were patients referred to the University Hospital Gasthuisberg Leuven and were of self-reported Flemish descent. Controls were recruited in the outpatient clinics in the University Hospital Leuven and were volunteers, free of neurological disease, accompanying patients. Scandinavian ALS cases and controls (comprising individuals from Sweden, Norway, Denmark and Finland) were collected as part of an ongoing prospective study on ALS genetics in the Nordic countries since 1993, with Umeå University in Sweden as the principal center. Controls were volunteers who accompanied patients, or unrelated, healthy individuals. Irish ALS cases were referred to Beaumont Hospital in Dublin, Ireland. Control subjects were spouses accompanying patients or unrelated volunteers and were recruited in the neurology outpatient clinic. German ALS cases were referred to the Department of Neurology of the University of Ulm and Charité University Hospital, Berlin. German controls were obtained from a genome-wide association study on schizophrenia and from a prospective cohort study for Parkinson’s disease in Tuebingen (40, 41). The study population from the United States consisted of 490 cases and 525 controls (11, 42). ALS cases were referred to the Massachusetts General Hospital, Boston or the Emory University Hospital, Atlanta. Controls were healthy volunteers and spouses of ALS patients from the Boston area. All cases and controls were of Western European descent. Polish cases were referred to the Motor Neuron Disease clinic of the Jagiellonian University, Krakow. Controls were from southern Poland, were free of neurological disease and without a family history of neurological diseases. Both cases and controls were of Polish ancestry.

Genotyping and CNV calling

All genome-wide datasets have been generated using DNA from fresh venous blood samples. All participants were genotyped on the Illumina platform according to the manufacturer’s protocol. The datasets were genotyped using different Illumina arrays: the HumanHap 300, HumanHap 550, HumanCNV 370 and the HumanHap 610. To avoid possible bias introduced by the different probe densities of the different arrays, we only used probes that were common to the four arrays (302,662 probes, essentially the content of the HumanHap 300 array).

Genome-wide data in the discovery and replication phase were handled as outlined in Figure S1. After scanning the arrays, data was uploaded and pre-processed using Illumina Beadstudio software. To remove artificial wave
artefacts, we applied loess correction to the log₂ R ratio, using Wavenorm (43). We then used PennCNV for CNV detection (44). We merged adjacent CNV calls of the same CNV type that were ≤ 10 probes and ≤ 50 kilobases apart. We then applied quality control filters, as described previously (45). We visually inspected each sample that produced ≥ 10 CNV calls and removed samples with obvious artifactual calls from further analyses.

**Quantitative PCR**

For the replication study we screened DNA samples with TaqMan quantitative PCR with subsequent validation. Quantitative PCR experiments were carried out on the Fluidigm platform using BioMark 96.96 Dynamic Arrays (Fluidigm Corporation) and TaqMan Copy Number Assays (Applied Biosystems). To analyse the two loci of interest, we used 12 target assays that were spatially distributed, so that each CNV detected in the genome-wide screen would be captured by at least one assay. As reference genes we used TaqMan Copy Number Reference assays, targeted to RNase P (Applied Biosystems), and to TIMM50 in the 19q13.2 locus. This locus does not contain known CNVs (Database of Genomic Variants) and the use of two instead of one reference assay reduced experimental noise. Assay details can be found in Table S10. Target and reference assays were run in separate PCR reactions and each reaction was done in triplicate. Each PCR plate included a no-template control and a standard curve of serially diluted DNA. Samples were first subjected to 14 cycles of Specific Target Amplification (STA) (46), using TaqMan PreAmp Mastermix (Applied Biosystems). Subsequent qPCR amplification was done using a standard TaqMan protocol.

We used Fluidigm Real-Time PCR Analysis software to determine raw Ct values, with default quality control thresholds. We then used qBase (47) software to calculate estimated copy numbers, corrected for plate- and assay-specific PCR efficiencies. We then filtered the resulting values as described previously (48): using the qcc package in R (49) we constructed c-charts to determine empirical cut-off values to identify copy number measurements with a high variability between replicates, or with a high variability between measurements calculated with different reference genes (indicating a failing PCR reaction for one of the two reference assays). Values that exceeded this cut-off were set to missing. Samples with a large number of failed assays (arbitrarily set to 80%) were excluded from further analyses. Putative CNVs were subsequently validated with TaqMan qPCR on an ABI Prism 7900HT Sequence detection System (Applied Biosystems) according to the
manufacturer’s protocol. Additionally, we examined the putative CNVs in all but four samples using the Illumina HumanOmni1-Quad arrays confirming their presence.

**Statistical analyses**

To test for differences in global CNV burden between cases and controls, we used Plink v1.06 software using 10,000 permutations to acquire empirical p-values. For association mapping we annotated CNVs with RefSeq gene IDs (downloaded from the UCSC genome browser, using the largest transcript per gene) and tested each gene for association of CNVs affecting the gene with ALS affection status. This approach has the advantage that CNVs with different breakpoints affecting the same gene are pooled into units with the same (assumed) functional impact, which provides more power than testing each variant individually (50). We tested each gene for association with all CNVs lumped, duplications only and deletions only. We used Fisher’s exact test and additionally tested associated loci with Cochran-Mantel-Haenszel test, stratified by country of origin. Because we only tested the hypothesis that variants were more prevalent in cases, we used one-sided tests. Genes with a p < 0.01 (one-sided Fisher’s exact test) and with a frequency of <1% in controls were selected for follow-up. Heterogeneity of the odds ratios between populations was evaluated using Woolf’s test in the vcd package in R. For stratification analyses we used the identity-by-state (IBS) analysis as implemented in Plink v1.06. We first generated 15,131 unlinked (r^2 < 0.50), non-disease-associated SNPs across the autosomal chromosomes with a call rate of 100% (4). We then calculated and plotted the first two dimensions of the multidimensional scaling analysis of IBS distances in all samples. We examined linkage disequilibrium properties of the **DPP6** locus in our study population using Haploview v4.2 (51). In order to calculate linkage disequilibrium we recoded duplications in the 5’ end of **DPP6** as one variant. Treating CNVs with different breakpoints separately did not produce meaningful results because of the low allele frequencies of these variants. Power calculations were done using G*Power software (52).
Acknowledgements

We would like to thank all patients and healthy volunteers participating in this study. This project was supported by the Prinse Beatrix Fonds, VSB fonds, H. Kersten and M. Kersten (Kersten Foundation), The Netherlands ALS Foundation, and J.R. van Dijk and the Adessium Foundation (L.H.v.d.B.). J.H.V. is supported by the Brain Foundation of The Netherlands. The GWA study was funded by the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project nr. 050-060-810. We thank Pascal Arp, Mila Jhamai, Dr Michael Moorhouse, Marijn Verkerk, and Sander Bervoets for their help in creating the GWAS database. The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are very grateful to the participants and staff from the Rotterdam Study, the participating general practitioners and the pharmacists. The Dutch controls from the Radboud University Nijmegen Medical Center (RUNMC) are part of the Nijmegen Biomedical Study, a survey from the general population of the municipality of Nijmegen. This study was conducted with an investment grant of the RUNMC. The Irish studies were funded by The Muscular Dystrophy Association (US), The Health Research Board of Ireland, The Irish Neurological Association Travel Award and The Irish Motor Neuron Disease Research Foundation.
References


### Summary of the study populations

<table>
<thead>
<tr>
<th>Platform</th>
<th>ALS cases</th>
<th>Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>272</td>
<td>276</td>
<td>Illumina CNV370</td>
</tr>
<tr>
<td>Ireland</td>
<td>204</td>
<td>205</td>
<td>Illumina 550K</td>
</tr>
<tr>
<td>Sweden</td>
<td>445</td>
<td>446</td>
<td>Illumina CNV370</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>954</td>
<td>7,804</td>
<td>Illumina 300K, CNV370, 550K</td>
</tr>
<tr>
<td>Total</td>
<td>1,875</td>
<td>8,731</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Platform</th>
<th>ALS cases</th>
<th>Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>1,032</td>
<td>994</td>
<td>Illumina 300K, Fluidigm</td>
</tr>
<tr>
<td>Ireland</td>
<td>100</td>
<td>127</td>
<td>Illumina 610K</td>
</tr>
<tr>
<td>Scandinavian</td>
<td>441</td>
<td>-</td>
<td>Fluidigm</td>
</tr>
<tr>
<td>Poland</td>
<td>217</td>
<td>372</td>
<td>Fluidigm</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>279</td>
<td>3,869</td>
<td>Illumina 550K, 610K, Fluidigm</td>
</tr>
<tr>
<td>United States</td>
<td>490</td>
<td>525</td>
<td>Illumina 300K</td>
</tr>
<tr>
<td>Total</td>
<td>2,559</td>
<td>5,887</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Platform</th>
<th>ALS cases</th>
<th>Controls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>4,434</td>
<td>14,618</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Results for $DPP6$ and 15q11.2 loci for genome-wide screen, replication and combined data.
Results for the $DPP6$ locus are for all CNVs (i.e. duplications and deletions) combined. For the 15q11.2 locus, results are for deletions only, as duplications did not contribute to the association signal. For a breakdown of results of the genome-wide screen please see Table S4.

<table>
<thead>
<tr>
<th>Locus</th>
<th>ALS cases</th>
<th>Controls</th>
<th>Fisher $P^a$</th>
<th>OR $^b$</th>
<th>CMH $P^c$</th>
<th>OR $^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Genome-wide screen:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7q36.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DPP6$</td>
<td>10 of 1,875</td>
<td>0.53</td>
<td>13 of 8,731</td>
<td>0.15</td>
<td>3.6x10$^{-3}$</td>
<td>3.59</td>
</tr>
<tr>
<td>15q11.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$TUBGCP5$</td>
<td>7 of 1,875</td>
<td>0.37</td>
<td>4 of 8,731</td>
<td>0.05</td>
<td>9.1x10$^{-4}$</td>
<td>8.17</td>
</tr>
<tr>
<td>$CYFIP1$</td>
<td>7 of 1,875</td>
<td>0.37</td>
<td>8 of 8,731</td>
<td>0.09</td>
<td>9.1x10$^{-3}$</td>
<td>4.09</td>
</tr>
<tr>
<td>$NIPA2$</td>
<td>8 of 1,875</td>
<td>0.43</td>
<td>7 of 8,731</td>
<td>0.08</td>
<td>1.9x10$^{-1}$</td>
<td>5.34</td>
</tr>
<tr>
<td>$NIPA1$</td>
<td>8 of 1,875</td>
<td>0.43</td>
<td>3 of 8,731</td>
<td>0.03</td>
<td>9.3x10$^{-3}$</td>
<td>12.46</td>
</tr>
<tr>
<td><strong>Replication:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7q36.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DPP6$</td>
<td>10 of 2,559</td>
<td>0.39</td>
<td>12 of 5,887</td>
<td>0.20</td>
<td>9.7x10$^{-2}$</td>
<td>1.92</td>
</tr>
<tr>
<td>15q11.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$TUBGCP5$</td>
<td>6 of 2,559</td>
<td>0.23</td>
<td>7 of 5,887</td>
<td>0.12</td>
<td>0.17</td>
<td>1.97</td>
</tr>
<tr>
<td>$CYFIP1$</td>
<td>6 of 2,559</td>
<td>0.23</td>
<td>7 of 5,887</td>
<td>0.12</td>
<td>0.17</td>
<td>1.97</td>
</tr>
<tr>
<td>$NIPA2$</td>
<td>7 of 2,559</td>
<td>0.27</td>
<td>6 of 5,887</td>
<td>0.10</td>
<td>4.2x10$^{-2}$</td>
<td>3.05</td>
</tr>
<tr>
<td>$NIPA1$</td>
<td>7 of 2,559</td>
<td>0.27</td>
<td>5 of 5,887</td>
<td>0.08</td>
<td>4.1x10$^{-2}$</td>
<td>3.23</td>
</tr>
<tr>
<td><strong>Combined:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7q36.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DPP6$</td>
<td>20 of 4,434</td>
<td>0.45</td>
<td>25 of 14,618</td>
<td>0.17</td>
<td>1.4x10$^{-3}$</td>
<td>2.64</td>
</tr>
<tr>
<td>15q11.2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$TUBGCP5$</td>
<td>13 of 4,434</td>
<td>0.29</td>
<td>11 of 14,618</td>
<td>0.08</td>
<td>1.0x10$^{-3}$</td>
<td>3.90</td>
</tr>
<tr>
<td>$CYFIP1$</td>
<td>13 of 4,434</td>
<td>0.29</td>
<td>15 of 14,618</td>
<td>0.10</td>
<td>5.9x10$^{-3}$</td>
<td>2.86</td>
</tr>
<tr>
<td>$NIPA2$</td>
<td>15 of 4,434</td>
<td>0.34</td>
<td>13 of 14,618</td>
<td>0.09</td>
<td>4.9x10$^{-4}$</td>
<td>3.81</td>
</tr>
<tr>
<td>$NIPA1$</td>
<td>15 of 4,434</td>
<td>0.34</td>
<td>8 of 14,618</td>
<td>0.05</td>
<td>2.2x10$^{-3}$</td>
<td>6.20</td>
</tr>
</tbody>
</table>

$^a$ One-tailed Fisher exact $p$ value

$^b$ Odds ratio

$^c$ Cochran-Mantel-Haenszel $p$ value (one-tailed), stratified by country of origin
Legends to figures:

**Figure 1.** Genomic organization of the *DPP6* and 15q11.2 loci

Figure shows duplications (green) and deletions (red) found in 1,875 ALS cases and 8,731 controls in the genome-wide screen. Segmental duplications are shown in yellow.