GENETIC RISK FACTORS FOR AMYOTROPHIC LATERAL SCLEROSIS

focusing on copy number variation

Genetische risicofactoren voor amyotrofische laterale sclerose,
in het bijzonder copy number variation

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 18 december 2012 des middags te 12.45 uur

door

Hylke Merijn Blauw

geboren op 26 juni 1979 te Eutin, Duitsland

Promotor: Prof. dr. L.H. van den Berg
Co-promotor: Dr. J.H. Veldink

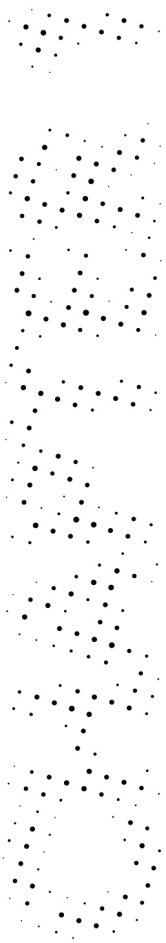
CONTENTS

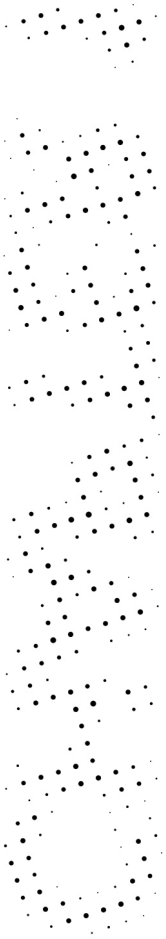
1	Introduction	5
2	*ITPR2* as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study	11
3	Genetic variation in *DPP6* is associated with susceptibility to amyotrophic lateral sclerosis	27
4	Genome-wide association study identifies 19p13.3 (*UNC13A*) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis	43
5	Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen	64
6	Analysis of genome-wide copy number variation in Irish and Dutch ALS populations	81
7	A large genome scan for rare CNVs in amyotrophic lateral sclerosis	97
8	Copy-number polymorphisms in amyotrophic lateral sclerosis	123
9	*NIPA1* polyalanine repeat expansions are associated with amyotrophic lateral sclerosis	137
10	*SMN1* gene duplications are associated with sporadic ALS	153
11	General discussion	163
	Nederlandse samenvatting	177
	Dankwoord	183
	List of publications	187
	Curriculum vitae	191

ISBN: 978-94-6191-526-9

English title: Genetic risk factors for amyotrophic lateral sclerosis; focusing on copy number variation
Nederlandse titel: Genetische risicofactoren voor amyotrofische laterale sclerose;
in het bijzonder copy number variation

Cover & book design: studio ilse van klei / www.ilsevanklei.nl





INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a uniformly fatal neurodegenerative disease affecting many patients worldwide for which there is no curative therapy. The disease is characterised by the selective death of motor neurons in the anterior horn of the spinal cord and motor cortex. The clinical symptoms are the result of upper and lower motor neuron death: muscle wasting, weakness, cramps, dysarthria and spasticity (1). ALS is a relentlessly progressive disease: progressive weakness usually culminates in neurogenic respiratory failure. Patients have a median survival of about three years from the onset of symptoms, while only a small percentage of patients have a slower disease progression and survive beyond 10 years (2, 3). Prognostic factors for a longer survival are a younger age at onset of the disease, spinal (versus bulbar) onset of symptoms and male gender (2). The only proven therapy that slows disease progression is with riluzole (Rilutek), which increases survival with about 3 months (4). ALS susceptibility is age-dependent: the incidence peaks in the sixth and seventh decade, and falls sharply thereafter (2). There is a male preponderance with a male to female ratio in incident cases of 1.3-1.5 (2, 5). The incidence of ALS is about 2-3 per 100,000 in Europe and the prevalence is about 10 per 100,000 (2, 5). The estimated lifetime risk is about 1/350-400 (6). So although not common, ALS is certainly not a rare disease.

The aetiology of ALS is largely unknown. Several hypotheses have been postulated about the pathogenic mechanisms that are operative in ALS. Oxidative stress, glutamate excitotoxicity, disrupted axonal transport, mitochondrial dysfunction, protein aggregation and altered RNA metabolism all seem to play a role in ALS pathogenesis (7). However, many of these hypotheses have been

inferred from animal models and monogenetic ALS in familial cases. To which degree these mechanisms can be translated to the "average" ALS patient and whether the proposed pathways represent up- or downstream effects is unclear.

Traditionally, ALS has been divided in familial and non-familial (sporadic) forms. About 5% of ALS cases are familial, although this number is probably an underestimate (8, 9). Several causative gene mutations have been identified by systematic analysis of pedigrees in linkage studies (using polymorphic markers to map disease-associated loci) and, more recently, in studies using whole genome sequencing technology. There are now over 15 different loci that harbour mutations that cause familial ALS (8, 10). The implicated genes include *SOD1, TARDP, VAPB, SETX, FUS, C9ORF72, ANG, OPTN and PFN1*. Some of these are associated with atypical forms with young onset, protracted disease course, with additional clinical features such as ataxia and parkinsonism, but most familial cases cannot be distinguished clinically from sporadic cases (8). Most pedigrees exhibit autosomal dominant modes of inheritance, although more complex patterns are frequently observed (8, 11).

The causes of sporadic ALS are essentially unknown. It is fair to say that there is not "one" cause for sporadic ALS. In contrast to pure monogenetic, familial ALS the pathogenesis of sporadic ALS is considered to be multifactorial in which an interplay of environmental and genetic factors contributes to disease liability. Environmental factors that have been proposed to influence the risk of ALS are smoking, diet, exposure to toxins and occupation, but for most the exact role is unclear (12-14).

There is increasing evidence that genetic factors play a major role in sporadic ALS, as well as in familial forms of ALS. In contrast to the monogenetic (Mendelian) basis for familial ALS, sporadic ALS is considered to be a complex genetic trait, in which multiple genetic variants predispose to disease by surpassing a liability threshold (8). The genetic basis is demonstrated by a relatively high heritability (the fraction of phenotypic variability of a trait that is explained by genetic variation). The heritability of sporadic ALS was estimated to be 0.61 (95% confidence interval: 0.38-0.78) in a large twin-study on ALS (15). There is an increased risk for relatives of people with ALS (16). Another indication of genetic contribution is the fact that the same mutations identified in ALS pedigrees have also been found in sporadic ALS cases (17, 18). Additionally, empirical support comes from findings in association studies in sporadic ALS, which show that genetic risk factors for sporadic ALS exist (19, 20). Therefore, studies dissecting the genetic basis for sporadic ALS can be a valuable instrument to gain insight into the biological pathways that underlie ALS susceptibility.

Despite the important genetic contribution to ALS pathogenesis, until recently, few well established associated genes had been identified. Many genes have been proposed as susceptibility genes, but most have not been replicated or showed ambiguous results (21). One example is the *SMN1* (Survival of Motor Neuron) gene. Because of its role in spinal muscular atrophy (SMA), a motor neuron disease with similarities to ALS, it was proposed as a candidate gene for ALS. A number of studies have reported associations with copy number differences in *SMN1* (and/ or its pseudogene SMN2), but because of conflicting results the exact mode of association was unclear (22-25). Similarly, for example, *VEGF* and *PON* have been proposed as ALS candidate genes, but their role in ALS remains controversial (21). More recently, however, a number of associations have been reported that convincingly identify *ATXN2, UNC13A* and *C9ORF72* as susceptibility genes for sporadic ALS (17-20, 26). These variants explain only a part of the phenotypic variance, indicating that most genetic factors are yet to be identified.

Recent technological developments have greatly facilitated the dissection of human complex traits and diseases. The completion of the Human Genome and HapMap

projects have boosted the knowledge of the human genome and genetic variation and have led to the development of new high-throughput SNP (single nucleotide polymorphisms) genotyping technology. Since only recently, SNP arrays can be used to rapidly genotype several hundreds of thousands marker in parallel in a great number of individuals. With this technology the entire genome can be screened for polymorphisms that associate to a certain phenotype in genome-wide association studies (GWAS) (27). This way associated loci can be searched for genome-wide, without any required hypothesis regarding the disease aetiology. This approach relies heavily on the "common disease-common variants" (CDCV) hypothesis, postulating that common (or complex) traits are influenced by (multiple) genetic variants, each with modest effects, occurring frequently (allele frequency typically >1-5%) in the population (28). Alternatively, the "multiple rare variants" hypothesis states that rare (e.g. <1% in the population) variants with a high effect size will contribute to disease. These hypotheses are not mutually exclusive and for most complex traits there will probably be a contribution of both common and rare alleles (28).

While the bulk of genetic variation was once thought to consist of single nucleotide variants, the advent of oligonucleotide array technology has uncovered an immense source of genetic variation, termed structural variation, which encompasses a spectrum of genomic alterations spanning kilobases to megabases. Structural variants are typically defined as genomic alterations that are > 1kb in length and include translocations, inversions, and copy number variants (CNVs) (29). The latter denote genomic segments that are variable in copy number, compared to a reference genome, which can be insertions, duplications or deletions (29). Classically, largely through observations in cytogenetic studies, large chromosomal structural alterations have been associated to severe developmental diseases, such as Down syndrome and Prader-Willi syndrome (30). However, with the use of oligonucleotide arrays with ever increasing resolution and whole genome sequencing it is now known that CNVs are abundant in the genome of healthy individuals (31, 32). More than 100 Mb (>3%) of the human genome consists of relatively common copy number variants and each genome will harbour more than a thousand CNVs (31). CNVs have a substantial biological impact: they can encompass genes and regulatory elements and may influence gene expression and (disease) phenotypes (33). Indeed, in addition to the known "genomic diseases" mentioned above, the number of more complex traits that have been associated with CNVs is growing (34-39).

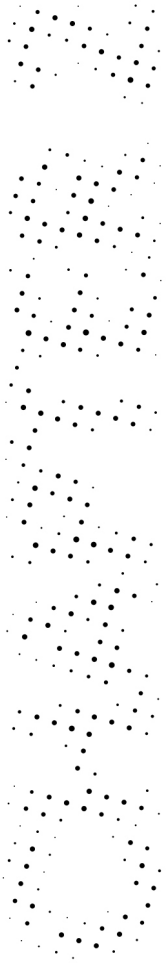
In conclusion, although sporadic ALS has a clear genetic basis, only little is known about the loci that confer susceptibility to ALS. Knowledge of these genetic risk factors and involved pathways will give greater insight into the involved pathobiology. This will facilitate the development of new treatment strategies, and possibly the identification of a susceptible subpopulation, so individuals at risk maybe treated with preventive strategies. The recent years have seen an explosion of knowledge and new technologies that have enabled the systematic dissection of the genetic basis of human traits and diseases. This thesis draws heavily on these newly available possibilities.

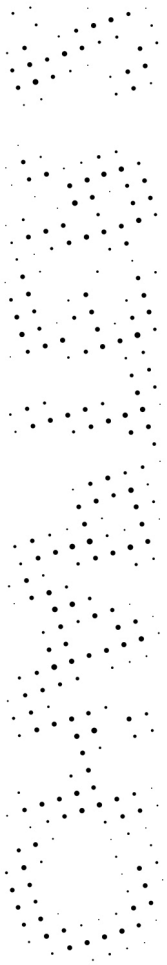
The aim of this thesis was to identify genetic risk factors that predispose to sporadic ALS and thus give new insights to biological mechanisms influencing ALS liability. We did this by 1) exploring common genetic variation in (SNP based) genome-wide association studies (Chapters 2,3,4), 2) exploring the role of rare and common CNVs (Chapters 5,6,7,8) and 3) characterising the role of two candidate genes (Chapters 9 and 10).

REFERENCES

1. Hardiman, O., van den Berg, L.H. and Kiernan, M.C. (2011) Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat Rev Neurol*, 7, 639-649.
2. Huisman, M.H.B., De Jong, S.W., Van Doormaal, P.T.C., Weinreich, S.S., Schelhaas, H.J., Van Der Kooi, A.J., De Visser, M., Veldink, J.H. and Van Den Berg, L.H. (2011) Population based epidemiology of amyotrophic lateral sclerosis using capture-recapture methodology. *Journal of Neurology, Neurosurgery & Psychiatry*, 82, 1165-1170.
3. Talbot, K. (2009) Motor neuron disease: the bare essentials. *Pract Neurol*, 9, 303-309.
4. Miller, R.G., Mitchell, J.D., Lyon, M. and Moore, D.H. (2007) Riluzole for amyotrophic lateral sclerosis (ALS)/ motor neuron disease (MND). *Cochrane Database Syst Rev*, CD001447.
5. Logroscino, G., Traynor, B.J., Hardiman, O., Chiò, A., Mitchell, D., Swingler, R.J., Millul, A., Benn, E., Beghi, E. and EURALS (2010) Incidence of amyotrophic lateral sclerosis in Europe. *J Neurol Neurosurg Psychiatr,* 81, 385-390.
6. Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R. and Zoing, M.C. (2011) Amyotrophic lateral sclerosis. *Lancet*, 377, 942-955.
7. Ferraiuolo, L., Kirby, J., Grierson, A.J., Sendtner, M. and Shaw, P.J. (2011) Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nature Reviews Neurology*, 7, 616-630.
8. Andersen, P.M. and Al-Chalabi, A. (2011) Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol,* 7, 603-615.
9. Byrne, S., Walsh, C., Lynch, C., Bede, P., Elamin, M., Kenna, K., Mclaughlin, R. and Hardiman, O. (2011) Rate of familial amyotrophic lateral sclerosis: a systematic review and meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry,* 82, 623-627.
10. Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., Mckenna-Yasek, D., Baron, D.M. et al. (2012) Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature*, 1-7.
11. Van Blitterswijk, M., Van Es, M.A., Hennekam, E.A.M., Dooijes, D., Van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., Van Der Kooi, A.J., De Visser, M. et al. (2012) Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet,* 21, 3776-3784.
12. de Jong, S.W., Huisman, M.H., Sutedja, N.A., van der Kooi, A.J., de Visser, M., Schelhaas, H.J., Fischer, K., Veldink, J.H. and van den Berg, L.H. (2012) Smoking, Alcohol Consumption, and the Risk of Amyotrophic Lateral Sclerosis: A Population-based Study. *Am J Epidemiol*, 176, 233-239.
13. Sutedja, N.A., Veldink, J.H., Fischer, K., Kromhout, H., Wokke, J.H., Huisman, M.H., Heederik, D.J. and Van den Berg, L.H. (2007) Lifetime occupation, education, smoking, and risk of ALS. *Neurology*, 69, 1508-1514.
14. Sutedja, N.A., Veldink, J.H., Fischer, K., Kromhout, H., Heederik, D., Huisman, M.H., Wokke, J.H. and van den Berg, L.H. (2009) Exposure to chemicals and metals and risk of amyotrophic lateral sclerosis: a systematic review. *Amyotroph Lateral Scler*, 10, 302-309.
15. Al-Chalabi, A., Fang, F., Hanby, M.F., Leigh, P.N., Shaw, C.E., Ye, W. and Rijsdijk, F. (2010) An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry,* 81, 1324-1326.
16. Hanby, M.F., Scott, K.M., Scotton, W., Wijesekera, L., Mole, T., Ellis, C.E., Leigh, P.N., Shaw, C.E. and Al-Chalabi, A. (2011) The risk to relatives of patients with sporadic amyotrophic lateral sclerosis. Brain, 134, 3454-3457.
17. Dejesus-Hernandez, M., Mackenzie, Ian R., Boeve, Bradley F., Boxer, Adam L., Baker, M., Rutherford, Nicola J., Nicholson, Alexandra M., Finch, NiCole A., Flynn, H., Adamson, J. et al. (2011) Expanded GGGGCC Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron*, 1-13.
18. Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., Van Swieten, J.C., Myllykangas, L. et al. (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, 72, 257-268.
19. Elden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M. et al. (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, 466, 1069-1075.
20. Van Es, M.A., Veldink, J.H., Saris, C.G.J., Blauw, H.M., Van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B. et al. (2009) Genome-wide association study identifies 19p13.3 (UNC13A) and

9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet*, 41, 1083-1087.
21. Schymick, J.C., Talbot, K. and Traynor, B.J. (2007) Genetics of sporadic amyotrophic lateral sclerosis. *Human Molecular Genetics*, 16, R233-R242.
22. Corcia, P., Camu, W., Halimi, J.-M., Vourc'h, P., Antar, C., Vedrine, S., Giraudeau, B., de Toffol, B., Andres, C.R. and Group, F.A.S. (2006) SMN1 gene, but not SMN2, is a risk factor for sporadic ALS. *Neurology*, 67, 1147-1150.
23. Corcia, P., Mayeux-Portas, V., Khoris, J., de Toffol, B., Autret, A., Müh, J.-P., Camu, W., Andres, C. and Sclerosis, F.A.R.G.A.L. (2002) Abnormal SMN1 gene copy number is a susceptibility factor for amyotrophic lateral sclerosis. *Ann Neurol.*, 51, 243-246.
24. Veldink, J.H., Kalmijn, S., Van der Hout, A.H., Lemmink, H.H., Groeneveld, G.J., Lummen, C., Scheffer, H., Wokke, J.H.J. and Van den Berg, L.H. (2005) SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. *Neurology*, 65, 820-825.
25. Veldink, J.H., Van den Berg, L.H., Cobben, J.M., Stulp, R.P., De Jong, J.M., Vogels, O.J., Baas, F., Wokke, J.H. and Scheffer, H. (2001) Homozygous deletion of the survival motor neuron 2 gene is a prognostic factor in sporadic ALS. *Neurology*, 56, 749-752.
26. Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M. et al. (2012) The C9ORF72 expansion mutation is a common cause of ALS +/-FTD in Europe and has a single founder. *European Journal of Human Genetics*, 1-7.
27. Pearson, T.A. and Manolio, T.A. (2008) How to Interpret a Genome-wide Association Study. *JAMA: The Journal of the American Medical Association*, 299, 1335-1344.
28. Gibson, G. (2012) Rare and common variants: twenty arguments. *Nat Rev Genet*, 13, 135-145.
29. Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. *Nat Rev Genet*, 7, 85-97.
30. Inoue, K. and Lupski, J.R. (2002) Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet*, 3, 199-242.
31. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P. *et al.* (2009) Origins and functional impact of copy number variation in the human genome. *Nature*, 1-9.
32. Mills, R.E., Walter, K., Stewart, C., Handsaker, R.E., Chen, K., Alkan, C., Abyzov, A., Yoon, S.C., Ye, K., Cheetham, R.K. et al. (2011) Mapping copy number variation by population-scale genome sequencing. *Nature*, 470, 59-65.
33. Hurles, M., Dermitzakis, E. and Tylersmith, C. (2008) The functional impact of structural variation in humans. *Trends in Genetics*, 24, 238-245.
34. De Cid, R., Riveira-Munoz, E., Zeeuwen, P.L.J.M., Robarge, J., Liao, W., Dannhauser, E.N., Giardina, E., Stuart, P.E., Nair, R., Helms, C. et al. (2009) Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet*, 41, 211-215.
35. Hollox, E.J., Huffmeier, U., Zeeuwen, P.L.J.M., Palla, R., Lascorz, J., Rodijk-Olthuis, D., Van De Kerkhof, P.C.M., Traupe, H., De Jongh, G., Heijer, M.D. et al. (2008) Psoriasis is associated with increased -defensin genomic copy number. *Nat Genet*, 40, 23-25.
36. De Kovel, C.G.F., Trucks, H., Helbig, I., Mefford, H.C., Baker, C., Leu, C., Kluck, C., Muhle, H., Von Spiczak, S., Ostertag, P. et al. (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain*, 133, 23-32.
37. Stone, J.L., O'donovan, M.C., Gurling, H., Kirov, G.K., Blackwood, D.H.R., Corvin, A., Craddock, N.J., Gill, M., Hultman, C.M., Lichtenstein, P. et al. (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455, 237-241.
38. Willer, C.J., Speliotes, E.K., Loos, R.J.F., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C. et al. (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*, 41, 25-34.
39. McCarroll, S.A., Huett, A., Kuballa, P., Chilewski, S.D., Landry, A., Goyette, P., Zody, M.C., Hall, J.L., Brant, S.R., Cho, J.H. et al. (2008) Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet*, 40, 1107-1112.





GENOME-WIDE ASSOCIATION STUDY OF SPORADIC AMYOTROPHIC LATERAL SCLEROSIS IDENTIFIES *ITPR2* AS A SUSCEPTIBILITY GENE

The Lancet Neurology (2007)

Hylke M Blauw[1,11*], Michael A van Es[1,11], Paul WJ van Vught[1,11], Lude Franke[2,11], Ludo Van Den Bosch[3], Sonja W de Jong[1], Vianney de Jong[4], Frank Baas[5], Ruben van 't Slot[2], Robin Lemmens[3], Helenius J Schelhaas[6], Anna Birve[7], Kristel Sleegers[8], Christine van Broeckhoven[8], John JH Wokke[1], Cisca Wijmenga[2], Wim Robberecht[3], Peter M Andersen[7], Jan H Veldink[1], Roel A Ophoff[9,10,12] & Leonard H. van den Berg[1,12]

[1] Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, the Netherlands; [2] Department of Medical Genetics, University Medical Center Utrecht, the Netherlands; [3] Department of Neurology, University Hospital Gasthuisberg, Leuven, Belgium; [4] Department of Neurology and [5] Neurogenetics, Academic Medical Center, Amsterdam, the Netherlands; [6] Department of Neurology, Radboud University Medical Center, Nijmegen, the Netherlands; [7] Institute of Clinical Neuroscience, Umeå University Hospital, Umeå, Sweden; [8] Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB, Antwerpen, Belgium; [9] Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, the Netherlands; [10] Neuropsychiatric Institute, University of California, Los Angeles, USA; [11] These authors contributed equally to this work; [12] L.H.vdB and RAO are lead investigators and contributed equally to this work

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a devastating disease characterized by progressive degeneration of motor neurons in the brain and spinal cord. It is considered to be a multi-factorial disease caused by environmental and genetic susceptibility factors. Recent technological developments have made it possible to perform genome-wide association studies, in which nearly all genetic variation is tested for association with disease in an unbiased manner.

We performed a genome-wide association study in 461 ALS-patients and 450 controls from The Netherlands using Illumina 300K SNP chips and replicated our findings in 2 independent sample series.

A genetic variant in the inositol 1,4,5-triphosphate receptor 2 *(ITPR2)* gene was found to be associated with ALS (P = 0.012 after Bonferroni correction). We replicated this finding in 2 independent samples. Combined analysis of all samples (total: 1,337 patients and 1,356 controls) gave a combined p-value of 3.28×10^{-6} and an odds ratio of 1.58 (95% CI 1.30-1.91). We found significantly elevated gene expression levels of *ITPR2* in peripheral blood of 126 ALS patients compared to 126 healthy controls (P = 0.00016).

We identified genetic variation in *ITPR2* as a susceptibility factor for ALS. *ITPR2* is an important regulator of intracellular Ca^{2+}-levels and is involved in glutamate-mediated neurotransmission of motor neurons. Elevated *ITPR2* gene expression levels, as found in our patients, are essential to apoptosis, which is the final common pathway in motor neuron death in ALS. Since *ITPR2* is involved with glutamate, Ca^{2+} and apoptosis, it is a strong biological candidate for a susceptibility gene in ALS.

Amyotrophic Lateral Sclerosis (ALS), also known as motor neuron disease, is a neurodegenerative disorder characterized by progressive wasting and weakness of limb, bulbar and respiratory muscles. It is caused by progressive loss of motor neurons in the spinal cord, brain stem and motor cortex and can occur at anytime in adulthood with a median age of onset in the mid-fifties. Approximately 50% of patients die within 3 years after onset of symptoms, usually as the result of respiratory failure.[1-3] Interference with glutamate-mediated toxicity via Riluzole is the only therapeutic strategy available to date slowing ALS disease progression by 3 to 6 months.[4]

Familial ALS accounts for approximately 10% of patients with 12 – 23% of families linked to mutations in *SOD1*.[5] Mutations in *Alsin*, *Dynactin*, *VAPB* and *ANG* have been found in rare cases of familial ALS.[6, 7] Sporadic ALS accounts for more than 90% of patients and is considered to be a multi-factorial disease, in which both environmental and genetic factors play a role.[8-10] The estimated heritability ranges from 0.38 – 0.85.[11]

The cause of motor neuron degeneration in sporadic ALS is not known, but hypotheses regarding ALS pathogenesis are many and include oxidative stress, glutamate excito-toxicity and apoptosis.[1, 12, 13] Based on these theories, multiple candidate gene studies have been performed, reporting associations with variants in multiple genes including *ANG*, *VEGF*, *HFE*, *PON1*, and copy number variations in *SMN1/SMN2*.[6, 14-20] In the candidate approach, a gene is selected based on prior knowledge of its function and subsequently tested for association with disease. Recently, technological and scientific progress has made it possible to identify genetic disease-susceptibility factors using a different approach. Genome-wide association studies allow researchers to screen nearly all common variation in the entire genome for association with disease in an unbiased way.[21, 22]

The vast majority of variation in human DNA is formed by Single Nucleotide

Polymorphisms (SNPs). To date, over 10,000,000 SNPs have been identified and it has been hypothesized that these common variants contribute to disease susceptibility (Common disease – common variant hypothesis). The International Haplotype Map project demonstrated that genetic variants that are near each other tend to be inherited together, meaning SNPs are inherited in groups or haplotypes. So-called "Tag SNPs" can be used to identify these haplotypes. Approximately 300,000 Tag SNPs contain most of the information about patterns of genetic variation for all 10 million common SNPs. The development of high density, high throughput genotyping technology has made it possible to perform genome-wide association studies, in which all Tag SNPs are tested for association with disease in one single experiment.[23]

To identify predisposing genetic factors for sporadic ALS, we performed a genome-wide association study in 461 patients and 450 healthy controls from The Netherlands and subsequently replicated our most significant findings in 2 independent sample series. Recent studies have made efforts to search for blood markers for diseases without obvious blood phenotypes. These studies have demonstrated that changes in gene-expression in peripheral blood reflect neurological diseases and have been demonstrated for cerebral infarction, Huntington's, Alzheimer's and Parkinson's disease.[24-27] To gain further insight into the possible pathological function of any associated gene, we generated gene-expression data from whole blood in 126 sporadic Dutch ALS patients and 126 age and gender matched healthy controls included in the genome-wide association study.

METHODS

Study design

This present study was designed according to a 3-stage genotyping design to reduce the burden for multiple testing and the chance of false positive associations.[28] In this design a very large number of SNPs is tested in the first stage and only the most significantly associated SNPs from stage I are selected for follow-up in subsequent stages. In all stages all patients and controls were genotyped individually.

In the first stage, over 300,000 SNPs were tested for association with ALS. The 500 most significantly associated SNPs were selected for further analysis in a second, independent population (stage II). In this study, all SNPs with $P < 0.1$ in stage II were selected for genotyping in a third stage. The stringency condition was set at the relatively low value of $P < 0.1$ to allow for genetic heterogeneity; Mutations in SOD1 are found in approximately 25% of families with ALS in the US, UK, Germany, Sweden and Belgium, but are very rare in Portugal, Switzerland and The Netherlands. Similar observations have been made for other genes including *ANG*, *VAPB* and *VEGF*. This design takes into account that ALS appears to be a genetically heterogeneous disease, even across European populations.[29]

Our study provided 80% power to detect an allelic association with an odds ratio ≥ 1.68, based on the observed average minor allele frequency of 26.1% and assuming the causal variant is typed or efficiently tagged. This study design effectively decreased the burden for multiple testing from 311,946 in the first genome-wide stage to 17 in the third stage of our study and thereby ensuring replication of associations in multiple populations (Figure 1).

To gain further insight into the possible pathological function of any associated gene, we simultaneously generated genome-wide gene-expression data using the 24k Illumina Sentrix HumanRef-8 Expression BeadChips on 126 sporadic Dutch ALS patients and 126 age and gender matched healthy controls included in the genome-wide association study.

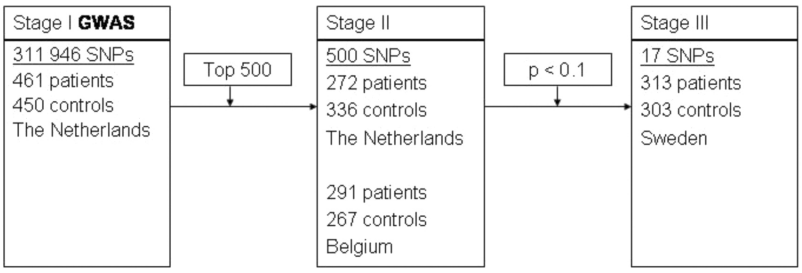


Figure 1 schematically shows the study design. In each stage a smaller number of SNPs is tested, thus decreasing the burden for multiple testing at each stage.

Study populations, sample collection & diagnosis

We analyzed populations from The Netherlands, Belgium and Sweden. In accordance with the principles of the Declaration of Helsinki (1964), with written informed consent and approved by the local ethical committees for medical research, venous blood samples were taken and DNA was extracted according to standard procedures. All samples from Belgian and Swedish patients were screened for *SOD1* gene mutations, the Swedish samples were also screened for ANG mutations. No samples with mutations in these genes were included in this study. Since no *SOD1* gene mutation has ever been reported in SALS or FALS in The Netherlands, no *SOD1* screening was done in these samples.[30]

All included patients were diagnosed by neurologists specialized in neuromuscular disease and ALS in particular. A detailed family history was taken from all included patients and no patients with a positive family history for ALS were included. All patients were diagnosed with sporadic ALS according to the 1994 El Escorial criteria.[31] Patients were included in the study when they fulfilled the criteria for probable or definite ALS.

All controls had negative medical and family histories for neurological disease; they were frequency-matched to cases for age, gender and ethnicity.

Dutch population: All included subjects were of Dutch ancestry (all 4 grandparents born in The Netherlands). All controls were frequency-matched to cases for age, gender and ethnicity. The 461 sporadic ALS patients included in the genome-wide association study were individuals referred to the University Medical Centre Utrecht (UMCU), the Academic Medical Centre Amsterdam (AMC) or the University Medical Centre Nijmegen, St Radboud. The 450 controls included in the genome-wide association study were unrelated, healthy volunteers accompanying non-ALS patients to the UMCU neurology out-patient clinic and spouses of sporadic ALS patients.

A total of 272 patients and 336 controls in the second, independent Dutch population were recruited from an ongoing, prospective population-based study on ALS in The Netherlands. In this study, a capture-recapture design is used to identify all prevalent and incident cases in The Netherlands.[32, 33] Family practitioners are then asked to recruit controls from their patient registers frequency matched to cases for age, gender and ethnicity.

Belgian population: Individuals with sporadic ALS were unrelated and, according to their own report, of Flemish descent for at least three generations. All patients were referred to the University Hospital Gasthuisberg, Leuven. The Belgian control group consisted of 267 unrelated, healthy, Flemish individuals that were selected from individuals who had married into families with neurological diseases collected for genetic studies.

Swedish population: The Swedish cohort consisted of 315 patients and 303 controls.

Individuals with sporadic ALS were unrelated Swedish citizens who reported to have had (northern) Swedish citizenship for at least three generations and were referred to the Umea University ALS Clinic. The Swedish control samples were spouses of the patients or unrelated healthy controls frequency-matched for age and gender.

Baseline characteristics for all populations are shown in Table 1.

Genotyping methods

In the first stage of this study, genotyping experiments were performed at the UMC Utrecht using the Illumina Infinium II HumanHap 300K SNP chips (Illumina, San Diego, CA, USA). All experiments were carried out according to the manufacturer's protocol. For the second stage, genotyping experiments were carried out using the Illumina GoldenGate assay (Illumina, San Diego, CA, USA). GoldenGate is a highly multiplexed PCR based genotyping assay. All experiments were carried out at the UMC Utrecht according to the manufacturer's protocol.

Table 1. Characteristics of study populations

Study Populations	Total (n)	Male / Female (%)	Spinal / Bulbar Onset (%)	Age at Onset* (years)	Survival** (months)
Genome wide study The Netherlands					
sALS patients	461	59 /41	69 / 31	59 (20-86)	33 (5-142)
Healthy Controls	450	59 / 41	-	60 (22-87)	-
Replication study The Netherlands					
sALS patients	272	57 / 43	71 / 29	58 (16-83)	43 (4-196)
Healthy Controls	336	57 / 43	-	59 (29-95)	-
Replication study Belgium					
sALS patients	291	59 / 41	73 / 27	59 (18-86)	39 (5-177)
Healthy Controls	267	58 / 42	-	51 (18-92)	-
Replication study Sweden					
sALS patients	313	57 / 43	66 / 34	60 (20-89)	31 (9-108)
Healthy Controls	303	52 / 48	-	62 (25-94)	-
Total					
sALS patients	1,337	58 / 42	70 / 30	59 (16-89)	36 (4-196)
Healthy Controls	1,356	57 / 43	-	58 (18-95)	-

*Age at onset shows median age in years with lowest and highest in parentheses for individuals with ALS and at blood sampling for healthy controls. **Survival shows median survival in months with lowest and highest value in parentheses from onset of initial weakness.

In stage III, SNPs were genotyped using Taqman assays. PCR was carried out with mixes consisting of 10 ng of genomic DNA, 1 x Taqman master mix (Applied Biosystems, Foster City, CA, USA), 1x assay mix (Applied Biosystems, Foster City, CA, USA) and ddH2O in a 5 µl reaction volume in 384-well plates (Applied Biosystems, Foster City, CA, USA). PCR conditions were as follows: denaturation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 92 °C for 15 seconds and annealing and extension at 60 °C for 1 minute.

Allelic PCR products were analyzed on the ABI Prism 7900HT Sequence Detection System using SDS 2.3 software (Applied Biosystems, Foster City, CA, USA). Primer and probe sequences are available upon request.

We randomly selected 100 individuals from stages I and II and genotyped them for all SNPs in stage III using the Taqman assays to ensure all platforms generated the same genotypes for each individual (concordance rate > 99.5%).

Detection of ITPR2 mRNA levels

After obtaining written, informed consent, DNA and RNA were collected from 126 sporadic Dutch ALS patients and 126 age- and gender-matched healthy controls. All blood samples were taken during the first visit to the neurology out-patient clinic in the UMCU. Samples were taken between 10AM and 12PM and before Riluzole treatment was initiated.

Blood for RNA isolation was drawn into PAXgene tubes (Qiagen, Valencia, CA, USA) and RNA was subsequently isolated using the PAXgene Blood RNA kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instruction, including an optional DNase digestion step. Total RNA was quantified using spectrophotometry and quality of total RNA was checked using Agilent 2 100 Bioanalyzer. *ITPR2* mRNA levels were measured using Illumina Sentrix humanref-8 arrays. 100 ng total RNA was used for first- and second-strand cDNA synthesis performed with a MessageAmp Kit (Ambion, Foster City, CA, USA). Hybridization and scanning of 24k Illumina Sentrix HumanRef-8 Expression BeadChips were performed according to the manufacturer's instructions (Illumina, San Diego, CA, USA). Data was obtained converting raw TIFF images into a data file using the Gene Expression Module of Illumina's BeadStudio software package. For normalization methods, the rank invariant algorithm was used.

Statistical analysis

We used PLINK 0.99r to perform the basic 1 df allelic chi-square tests for each SNP.[34] Odds ratios with 95% CI were calculated for the minor allele of each SNP. All SNPs were tested for deviations from Hardy-Weinberg equilibrium. We used Eigenstrat to detect evidence of population stratification.[35] Using SPSS we performed a Shapiro-Wilk W test of normality on the expression data, which showed that expression levels did not follow a normal distribution ($P < 0.001$). To compare *ITPR2* expression between cases and controls, a Wilcoxon-Mann-Whitney U test was therefore performed.

RESULTS

In the genome-wide study, we genotyped 477 unique samples from Dutch sporadic ALS patients and 472 unique Dutch controls. We excluded 34 individuals (12 patients and 22 controls) for poor quality genotyping (call rate <95%). Two pairs of patients (n=4) were excluded due to observed family relationships (>200,000 concordant SNPs). Two samples were genotyped twice yielding concordance of over 99.9% for each sample. The average call rate across all samples was 99.5%. The call rate was ≥99% for 298,807 SNPs and ≥95% for 315,293 SNPs. HWE was calculated in controls for each SNP and ≤0.05 for 14,498 SNPs

and ≤0.01 for 3,207 SNPs. The average minor allele frequency was 26·1%. 396 SNPs had a minor allele frequency <0.01. After pruning SNPs according to frequency (MAF <0.01) and genotyping (missingness ≥0.05, HWE ≤0.01), 311,946 SNPs remained. A total of 284,182,806 unique genotype calls were made.

The basic 1 df allelic chi-square tests was performed on all SNPs with minor allele frequency (MAF) ≥0.01, call rate ≥95%, and Hardy-Weinberg equilibrium $P \geq 0.01$. Results for all 311,946 SNPs analyzed in stage 1 can be found online at (http://www.alscentrum.nl/index.php?id=GWA.)

After applying the same quality control in stage II, we included 272 patients and 336 controls from The Netherlands and 291 patients and 267 controls from Belgium. Of the 500 selected SNPs, 485 were included for the analysis in stage II (call rate > 95% and MAF >0.01). Average call rate was 98.2% and average MAF was 23.5%. Eigenstrat analysis between Dutch and Belgian controls showed no evidence of population stratification and so we analyzed the Dutch and Belgian populations as one group. Results for all SNPs in stage II are shown in Table 2.

A total of 17 SNPs revealed a p-value <0.1 with the same direction of allele frequency and were subsequently selected for genotyping in a population of 313 patients and 303 controls from Sweden. Results for all SNPs tested in the third stage are shown in Table 3. Stage III analysis revealed significant association with sporadic ALS for one SNP, rs2306677 at P = 0.0007 (P = 0.012 after Bonferroni correction for testing of 17 SNPs). Considering that low minor allele frequencies are a potential form of bias in genetic association studies, we recalculated the statistics for rs2306677 using Fisher's Exact Test, which confirmed the initial results.

Table 2. SNPs with p<0.1 in Stage II are shown.

db SNP ID	Chr	Chr Pos	MAF*		HWE**		Allelic P
			Controls	Patients	Controls	Patients	
rs1929492	9	103,206,143	0.06	0.1	0.16	0.01	0.004
rs280199	2	121,449,946	0.29	0.24	0.84	0.65	0.013
rs2837501	21	40,499,803	0.45	0.5	0.56	0.67	0.013
rs4551564	1	222,304,339	0.42	0.37	0.5	0.28	0.027
rs721363	3	113,659,820	0.2	0.24	0.38	0.82	0.035
rs4620270	8	125,506,992	0.13	0.16	0.28	0.76	0.04
rs6559732	9	85,358,512	0.13	0.16	0.85	0.11	0.044
rs543721	2	136,878,027	0.41	0.37	0.31	0.53	0.048
rs1565730	12	65,143,477	0.26	0.29	1	0.54	0.051
rs3109032	3	113,656,323	0.2	0.24	0.32	0.82	0.054
rs6948572	7	137,312,129	0.2	0.23	0.44	0.55	0.069
rs1007241	13	89,136,899	0.37	0.41	0.19	0.22	0.083
rs9257425	6	29,044,535	0.26	0.29	0.25	0.92	0.084
rs901709	23	25,800,777	0.26	0.22	0.33	0.06	0.088
rs3026935	1	157,384,314	0.08	0.06	0.76	0.71	0.089
rs2306677	12	26,527,653	0.08	0.11	0.42	0.37	0.09
rs30264	5	129,514,804	0.1	0.13	0.51	0.85	0.098

*Minor Allele Frequency. **Hardy-Weinberg Equilibrium

Table 3. Stage 3 results

db SNP ID	Chr	Chr Pos	MAF*		HWE**		Allelic P
			Controls	Patients	Controls	Patients	
rs2306677	12	26,527,653	0.05	0.1	0.36	0.56	0.0007
rs30264	5	129,514,804	0.1	0.13	0.51	0.85	0.03
rs6559732	9	85,358,512	0.13	0.16	0.85	0.11	0.07
rs9257425	6	29,044,535	0.26	0.29	0.25	0.92	0.08
rs1565730	12	65,143,477	0.26	0.29	1	0.54	0.14
rs3109032	3	113,656,323	0.2	0.24	0.32	0.82	0.34
rs721363	3	113,659,820	0.2	0.24	0.38	0.82	0.37
rs3026935	1	157,384,314	0.08	0.06	0.76	0.71	0.44
rs4620270	8	125,506,992	0.13	0.16	0.28	0.76	0.56
rs2837501	21	40,499,803	0.45	0.5	0.56	0.67	0.58
rs901709	23	25,800,777	0.26	0.22	0.33	0.06	0.63
rs280199	2	121,449,946	0.29	0.24	0.84	0.65	0.79
rs4551564	1	222,304,339	0.42	0.37	0.5	0.28	0.83
rs543721	2	136,878,027	0.41	0.37	0.31	0.53	0.85
rs1007241	13	89,136,899	0.37	0.41	0.19	0.22	0.88
rs1929492	9	103,206,143	0.06	0.1	0.16	1	0.95
rs6948572	7	137,312,129	0.2	0.23	0.44	0.55	0.98

In stage III, 17 SNPs were tested for association. Allelic p-values for all SNP are shown.
*Minor Allele Frequency. **Hardy-Weinberg Equilibrium.

Table 4. Statistics for rs2306677 per population

Population	MAF*		HWE**		Allelic P	Fisher's Exact P	OR (95% CI)
	Controls	Patients	Controls	Patients			
The Netherlands (Stage I)	0.06	0.11	0.40	0.80	0.0005	0.0005	1.81 (1.29-2.55)
Belgium (Stage II)	0.10	0.09	0.50	0.47	0.413	0.4159	0.85 (0.57-1.26)
The Netherlands (Stage II)	0.07	0.12	0.67	0.57	0.0017	0.0015	1.85 (1.25-2.73)
Sweden (Stage III)	0.05	0.10	0.36	0.56	0.0007	0.0009	2.12 (1.35-3.33)
Netherlands pooled	0.06	0.11	0.51	0.62	3.73×10^{-6}	3.58×10^{-6}	1.81 (1.41-2.35)
Overall pooled	0.07	0.11	0.47	0.61	3.28×10^{-6}	3.76×10^{-6}	1.58 (1.30-1.91)

*Minor allele frequency. **Hardy-Weinberg Equilibrium p-values.
P-values were calculated using both the basic allele test and Fisher's Exact Test.

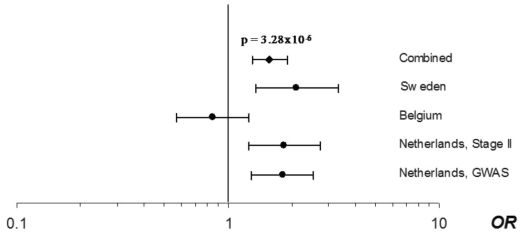


Figure 2 *shows the Odds Ratios for rs2306677 in each population and an Odds Ratio for all populations (sample series) combined. Error bars show 95% confidence intervals.*

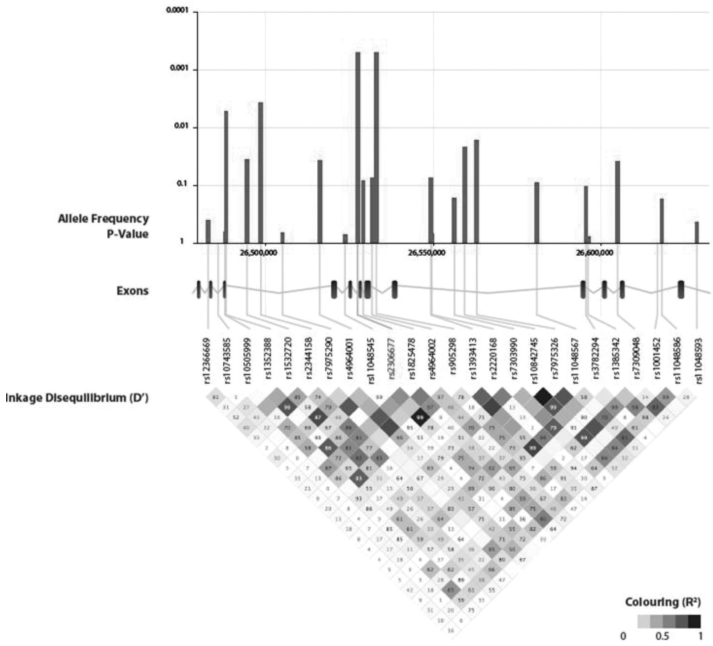
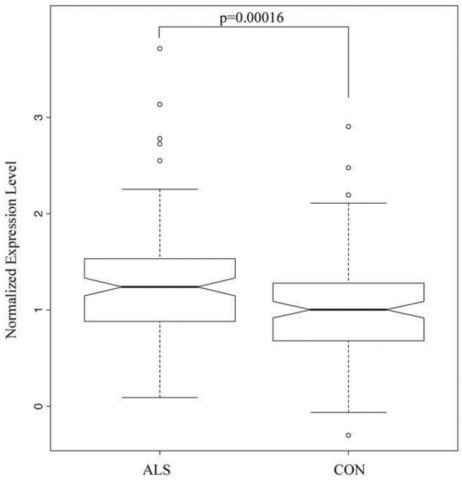


Figure 3 *shows a schematic view of association results, marker density, and Linkage Disequilibrium (LD) structure for the associated locus containing rs2306677 within ITPR2 in the Dutch GWA population.*
Top part of figure shows allelic p-values for each SNP. Bottom part shows LD pattern. LD is given in r^2 and is scored on a scale between 0 – 1. A high value for r^2 (>0·8) between two SNPs indicates that these genetic variants are inherited together more often than can be accounted for by chance and generally speaking indicate that the two alleles are physically close to each other on the DNA strand. Whereas lower values indicate that inheritance of genetic variants is independent and that they are likely not in close proximity. Values for r^2-values between SNPs surrounding rs2306677 are shown in the boxes in the LD plot, for instance the r^2-value between rs2306677 and rs905298 is 0·99 (by convention 0· is not shown in LD plots). These SNPs show similar p-values, which is to be expected as they tend to be inherited together. Values for r^2-values between SNPs are shown for a 150 kb region. Using an r^2> value of 0·8 a 45 kb LD block containing rs2306677 can be defined, which spans a region from rs10505999 in intron 39 to rs905298 in intron 45 within ITPR2. From this one can conclude that the associated variant is located within this 45 kb block and is rs2306677 or in close proximity to it.



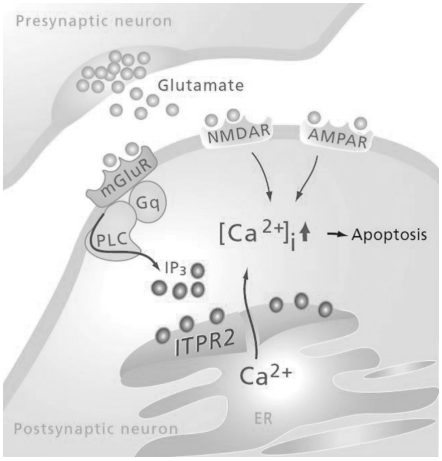


Figure 4 (left) *shows a boxplot of normalized expression levels for ITPR2 in 126 ALS patients and 126 healthy matched controls from the initial GWA study with median value (horizontal bar), 25th-75th inter-quartile range (box), maximum and minimum values, and outliers. Expression levels are significantly elevated in ALS patients (P = 0.00016).*

Figure 5 (right) *shows the proposed pathway whereby ITPR2 contributes to cell death. After stimulation of the metabotropic glutamate receptor (mGluR), inositol 1,4,5-triphosphate (IP3) is released. IP3 then binds its receptor, the inositol 1,4,5-triphosphate receptor type 2 (ITPR2), which results in release of Ca2+ from the endoplasmatic reticulum (ER) into the cytosol. IP3 levels in combination with raised ITPR2 levels result in elevated concentration of intracellular Ca2+, leading to apoptosis*

Overall analysis of the combined sample of Stages I-III, consisting of 1,337 patients and 1,356 controls yielded a p-value of 3.28×10^{-6} and an odds ratio of 1.58 (95% CI: 1.30-1.91). Table 4 shows the p-values for each individual population. Figure 2 shows the odds ratio for each population.

Rs2306677 is located in intron 42 of a gene encoding the inositol 1,4,5-phosphate receptor type 2 (*ITPR2*). Since the association signal maps to a 45 kb LD block (r^2 >0.6) within the gene (Figure 3) we consider *ITPR2* to be the ALS-associated gene.
We, therefore, investigated *ITPR2* gene expression levels in peripheral blood of 126 Dutch sporadic ALS patients and 126 matched controls available from Stage I. We observed significant, elevated gene expression levels in patients compared to controls (*P* = 0.00016, Figure 4). Unfortunately, it was not possible to analyze the correlation between genotype and expression, because RNA was not available from a sufficient number of patients and controls who were homozygous for the risk allele.

DISCUSSION

The completion of the HapMap-project and development of high through-put, high density, genotyping technology have made genome-wide association studies feasible. This approach enables the entire genome to be screened for association with disease

in an unbiased manner, in contrast to the candidate gene approach which is based on pathophysiological hypotheses. Indeed, over the last year, numerous genome-wide association studies have successfully identified novel genetic risk factors for multiple diseases, such as age-related Macula Degeneration, Diabetes Mellitus types 1 & 2, celiac disease and breast cancer.36-39

In this study we performed a genome-wide association study according to a 3-stage genotyping design. This design is an effective strategy for dealing with the over-conservative Bonferroni correction and reduced the burden for multiple testing from 311,924 (in stage I) to 17 (in stage III).

We found one SNP, rs2306677, to be significantly associated with ALS after correction for multiple testing (corrected $P = 0.012$). Combined analysis of 1,337 patients and 1,356 controls yielded a p-value of 3.28×10^{-6} and an odds ratio of 1.58 (95% CI: 1.30-1.91). Re-evaluation of Stage II results for SNP rs2306677 by population showed contrasting results for the Dutch and Belgian populations, with $P = 0.0017$ and $P = 0.41$, respectively. This observation suggests a consistent pattern of association of rs2306677 with ALS in the two independent Dutch and Swedish sample series. The lack of replication in the Belgian sample may be due to several factors, including the relatively small size of the sample resulting in inaccurate estimates of the allele frequencies, type I error, population specific environmental factors and the genetic heterogeneity of the disease even across European populations.[29] For instance, sequence variations in *ANG* are associated with ALS in Ireland and Scotland, but not in English, Swedish or Italian populations.[6] Previous studies have reported discrepancies in genetic associations between the Dutch and Belgian populations for ALS. Mutations in SOD1 are found in approximately 25% of families with ALS in Belgium, but are not found in The Netherlands.[30] Similarly, genetic variation in the promoter region of *VEGF* has been found to be associated with sporadic ALS in Belgium, but not in The Netherlands.[40]

Earlier this year first-stage data was released of a genome-wide association study of 271 ALS cases and 276 controls from the US. No genome-wide significant findings were observed, most likely due to the small sample size and thus limited power. This study had 80% power to detect an odds ratio of 2.75 and therefore failed to detect the association for rs2306677.[41] However, analysis of rs2306677 revealed a p-value of 0.14 in this population. Adding this SNP data to our study would further support association between ALS and rs2306677 ($P = 1.48 \times 10^{-6}$ versus $P = 3.28 \times 10^{-6}$ without US data; minor allele frequency of 0.12 and 0.09 for patients and controls, respectively).

Recently, a second genome-wide study of ALS was completed.[42] This study used a design that was similar to ours; a large number of SNPs were screened in the initial phase and a smaller number of candidate SNPs are carried on in following phases, hereby decreasing the burden for multiple testing. In this study the top 384 SNPs, and not all SNPs with $p<0{\cdot}05$, were selected for further analysis. Results for all SNPs from this second genome-wide study are not publicly available. Therefore, unfortunately no data on SNPs in *ITPR2* is available from this study.

Rs2306677 is located in intron 42 of a gene on chromosome 12p11 encoding the Inositol 1,4,5-triphosphate receptor type 2 (*ITPR2*). Since the association signal maps to a 45 kb LD block containing 5 exons ($r^2 > 0.6$) within the gene, we consider *ITPR2* to be the ALS- associated gene (Figure 3). *ITPR2* is a Ca^{2+}-release channel located on the endoplasmatic reticulum and is one of the proteins primarily responsible for controlling intracellular Ca^{2+}-levels in neurons (Figure 5). *ITPR2* is part of the Inositol 1,4,5-triphosphate receptor (ITPR) family consisting of three members (type 1-3). In the nervous system *ITPR2* is highly expressed in motor neurons.[43]

In motor neurons, *ITPR2* is bound by Inositol 1,4,5-triphosphate (IP3), which is

released after stimulation of glutamate receptors, and controls subsequent changes in intracellular Ca^{2+}-levels.[44] Altered function of *ITPR2* may lead to increased vulnerability to high intracellular Ca^{2+} levels, subsequently leading to apoptosis and hereby selective degeneration of motor neurons, the hallmark of ALS[1]. Evidence indicates that Ca^{2+} release from the endoplasmatic reticulum by *ITPR2* is crucial to both the extrinsic death receptor pathway and the intrinsic mitochondrial / cytochrome C apoptosis pathway.[45,46-48] Genetically altered cells lacking *ITPR2* demonstrate defect apoptosis, while cells over-expressing *ITPR2* show increased cell death.[49] Apoptosis is considered to be the final common pathway leading to motor neuron degeneration in ALS.[12] Investigation of *ITPR2* expression levels in peripheral blood of 126 Dutch sporadic ALS patients and 126 matched controls available from Stage I revealed significantly, elevated gene expression levels in patients compared to controls. The increased mRNA expression levels in peripheral blood suggest that multiple cell types of ALS patients are more susceptible to apoptosis[45], in particular cells with limited Ca^{2+}-buffering capacity such as motor neurons.[50] Recently it has been demonstrated that disease specific gene-expression profiles in peripheral blood can be demonstrated for neurological conditions such as cerebral infarction, Huntington's, Parkinson's and Alzheimer's disease.[24-27]
The power of genome-wide studies is increasingly recognized. However, the high sensitivity of genome-wide association studies means that there will probably be false-positive loci. Therefore, replication of associations in well designed, adequately powered studies is essential.

On the other hand, methods used for correcting for multiple testing, such as Bonferroni, tend to lead to type 2 errors. Although the Bonferroni method is appropriate for epidemiological studies investigating a few independent variables, disease-associated SNPs are unlikely to operate independently of each other and thus this approach will probably overcorrect.[8] This point is demonstrated by the fact that associations, which have been replicated in multiple populations, such as HFE and PON, have not detected by the genome-wide association studies in ALS to date.

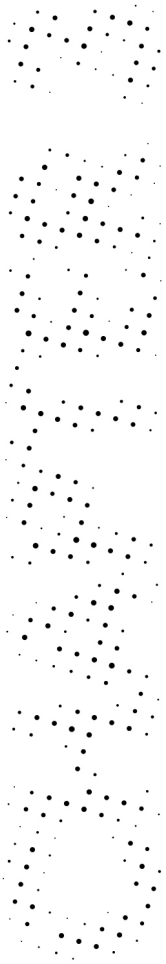
Due to the sample size of the population analyzed in Stage I and the limited number of SNPs carried forward (500), this study is not a completely comprehensive screen for ALS-susceptibility factors of the genome. It is, therefore, likely that larger or collaborative genome-wide efforts will reveal novel genetic susceptibility factors for ALS, considering this study lacked adequate power to detect odds ratio smaller than 1.68.
Nonetheless, this present study is the largest study, to date, and provides evidence for a novel ALS-susceptibility gene with plausible biological evidence for its role in disease pathogenesis

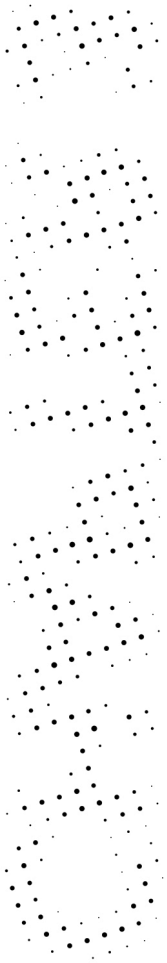
In summary, our genome-wide association study has identified *ITPR2* as susceptibility gene for sporadic ALS. The identification of *ITPR2* highlights the discovery of a biological highly relevant gene in the pathogenesis of ALS, considering (1) *ITPR2* is highly expressed in motor neurons compared to other neurons, (2) plays a role in glutamate mediated neurotransmission, (3) is the main regulator of intracellular Ca^{2+}-levels, (4) over-expression is essential to apoptosis and (5) ALS patients over-express *ITPR2*. Identification of a common variant within *ITPR2* is the first step of many in the genetic study of sporadic ALS, and it opens up new avenues for studying the molecular basis of this devastating disease.

REFERENCES

1. Cleveland,D.W. & Rothstein,J.D. From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. *Nat. Rev. Neurosci.* 2, 806-819 (2001).
2. Mitchell,J.D. & Borasio,G.D. Amyotrophic lateral sclerosis. *Lancet* 369, 2031-2041 (2007).
3. Rowland,L.P. & Shneider,N.A. Amyotrophic lateral sclerosis. *N. Engl. J. Med.* 344, 1688-1700 (2001).
4. Lacomblez,L., Bensimon,G., Leigh,P.N., Guillet,P., & Meininger,V. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. *Lancet* 347, 1425-1431 (1996).
5. Rosen,D.R. *et al*. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362, 59-62 (1993).
6. Greenway,M.J. *et al*. ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat. Genet.* 38, 411-413 (2006).
7. Nishimura,A.L. *et al*. A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am. J. Hum. Genet.* 75, 822-831 (2004).
8. Hardiman,O. & Greenway,M. The complex genetics of amyotrophic lateral sclerosis. *Lancet Neurol.* 6, 291-292 (2007).
9. Simpson,C.L. & Al-Chalabi,A. Amyotrophic lateral sclerosis as a complex genetic disease. *Biochim. Biophys. Acta* 1762, 973-985 (2006).
10. Veldink,J.H., Van den Berg,L.H., & Wokke,J.H. The future of motor neuron disease: the challenge is in the genes. *J. Neurol.* 251, 491-500 (2004).
11. Graham,A.J., Macdonald,A.M., & Hawkes,C.H. British motor neuron disease twin study. *J. Neurol. Neurosurg. Psychiatry* 62, 562-569 (1997).
12. Sathasivam,S., Ince,P.G., & Shaw,P.J. Apoptosis in amyotrophic lateral sclerosis: a review of the evidence. *Neuropathol. Appl. Neurobiol.* 27, 257-274 (2001).
13. Van Den,B.L., Van,D.P., Bogaert,E., & Robberecht,W. The role of excitotoxicity in the pathogenesis of amyotrophic lateral sclerosis. *Biochim. Biophys. Acta* 1762, 1068-1082 (2006).
14. Frankish,H. VEGF implicated in degeneration of motor neurons. *Lancet* 357, 1856 (2001).
15. Goodall,E.F. *et al*. Association of the H63D polymorphism in the hemochromatosis gene with sporadic ALS. *Neurology* 65, 934-937 (2005).
16. Kasperaviciute,D. *et al*. Large-scale pathways-based association study in amyotrophic lateral sclerosis. *Brain* (2007).
17. Lambrechts,D. *et al*. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nat. Genet.* 34, 383-394 (2003).
18. Saeed,M. *et al*. Paraoxonase cluster polymorphisms are associated with sporadic ALS. *Neurology* 67, 771-776 (2006).
19. Sutedja,N.A. *et al*. The association between H63D mutations in HFE and amyotrophic lateral sclerosis in a Dutch population. *Arch. Neurol.* 64, 63-67 (2007).
20. Veldink,J.H. *et al*. SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. *Neurology* 65, 820-825 (2005).
21. Davey,S.G. *et al*. Genetic epidemiology and public health: hope, hype, and future prospects. *Lancet* 366, 1484-1498 (2005).
22. Cordell,H.J. & Clayton,D.G. Genetic association studies. *Lancet* 366, 1121-1131 (2005).
23. The International HapMap Project. *Nature* 426, 789-796 (2003).
24. Borovecki,F. *et al*. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. *Proc. Natl. Acad. Sci. U.S.A.* 102, 11023-11028 (2005).
25. Tang,Y., Lu,A., Aronow,B.J., & Sharp,F.R. Blood genomic responses differ after stroke, seizures, hypoglycemia, and hypoxia: blood genomic fingerprints of disease. *Ann. Neurol.* 50, 699-707 (2001).
26. Maes,O.C. *et al*. Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. *Neurobiol. Aging* (2006).
27. Scherzer,C.R. *et al*. Molecular markers of early Parkinson's disease based on gene expression in blood. *Proc. Natl. Acad. Sci. U.S.A.* 104, 955-960 (2007).

28. Wen,S.H., Tzeng,J.Y., Kao,J.T., & Hsiao,C.K. A two-stage design for multiple testing in large-scale association studies. *J. Hum. Genet.* 51, 523-532 (2006).
29. Cronin,S., Hardiman,O., & Traynor,B.J. Ethnic variation in the incidence of ALS: a systematic review. *Neurology* 68, 1002-1007 (2007).
30. Baas,F. & Andersen,P.M. 2007. (Personal Communication)
31. Brooks,B.R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. *J Neurol Sci* 124 Suppl, 96-107 (1994).
32. Chao,A., Tsay,P.K., Lin,S.H., Shau,W.Y., & Chao,D.Y. The applications of capture-recapture models to epidemiological data. *Stat. Med.* 20, 3123-3157 (2001).
33. Hook,E.B. & Regal,R.R. Capture-recapture methods. *Lancet* 339, 742 (1992).
34. Dorus,S. *et al.* Accelerated evolution of nervous system genes in the origin of Homo sapiens. *Cell* 119, 1027-1040 (2004).
35. Price,A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904-909 (2006).
36. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661-678 (2007).
37. Easton,D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* (2007).
38. Klein,R.J. *et al.* Complement factor H polymorphism in age-related macular degeneration. *Science* 308, 385-389 (2005).
39. van Heel,D.A. *et al.* A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat. Genet.* (2007).
40. Van Vught,P.W. *et al.* Lack of association between VEGF polymorphisms and ALS in a Dutch population. *Neurology.* 65, 1643-1645 (2005).
41. Schymick,J.C. *et al.* Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 6, 322-328 (2007).
42. Dunckley,T. *et al.* Whole-Genome Analysis of Sporadic Amyotrophic Lateral Sclerosis. *N. Engl. J. Med.* (2007).
43. Yamamoto-Hino,M. *et al.* Cloning and characterization of human type 2 and type 3 inositol 1,4,5-trisphosphate receptors. *Receptors. Channels.* 2, 9-22 (1994).
44. Mikoshiba,K. Inositol 1,4,5-trisphosphate IP(3) receptors and their role in neuronal cell function. *J. Neurochem.* 97, 1627-1633 (2006).
45. Choe,C.U. & Ehrlich,B.E. The inositol 1,4,5-trisphosphate receptor (IP3R) and its regulators: sometimes good and sometimes bad teamwork. *Sci. STKE.* 2006, re15 (2006).
46. Boehning,D., Patterson,R.L., & Snyder,S.H. Apoptosis and calcium: new roles for cytochrome c and inositol 1,4,5-trisphosphate. *Cell Cycle 3,* 252-254 (2004).
47. Tantral,L. *et al.* Intracellular calcium release is required for caspase-3 and -9 activation. *Cell Biochem. Funct.* 22, 35-40 (2004).
48. Wozniak,A.L. *et al.* Requirement of biphasic calcium release from the endoplasmic reticulum for Fas-mediated apoptosis. *J. Cell Biol.* 175, 709-714 (2006).
49. Gutstein,D.E. & Marks,A.R. Role of inositol 1,4,5-trisphosphate receptors in regulating apoptotic signaling and heart failure. *Heart Vessels* Suppl 12, 53-57 (1997).
50. Van Den,B.L. *et al.* Calcium handling proteins in isolated spinal motoneurons. *Life Sci.* 65, 1597-1606 (1999).





GENETIC VARIATION IN *DPP6* IS ASSOCIATED WITH SUSCEPTIBILITY TO AMYOTROPHIC LATERAL SCLEROSIS

Nature Genetics (2008)

Hylke M Blauw[1,15], Michael A van Es[1,15], Paul WJ van Vught[1,15], Lude Franke[2,15], Christiaan GJ Saris[1], Ludo Van Den Bosch[3], Sonja W de Jong[1], Vianney de Jong[4], Frank Baas[5], Ruben van 't Slot[2], Robin Lemmens[3], Helenius J Schelhaas[6], Anna Birve[7], Kristel Sleegers[8,9], Christine Van Broeckhoven[8,9], Jennifer C Schymick[10], Bryan J Traynor[11], John HJ Wokke[1], Cisca Wijmenga[2,12], Wim Robberecht[3], Peter M Andersen[7], Jan H Veldink[1], Roel A Ophoff[13,14,16] & Leonard H. van den Berg[1,16]

[1]Department of Neurology, Rudolf Magnus Institute of Neuroscience and [2]Complex Genetics Section, Department of Biomedical Genetics, University Medical Center Utrecht, The Netherlands; [3]Department of Neurology, University Hospital Gasthuisberg, Leuven, Belgium; [4]Department of Neurology and [5]Neurogenetics, Academic Medical Center, Amsterdam, The Netherlands; [6]Department of Neurology, Radboud University Medical Center, Nijmegen, The Netherlands; [7]Institute of Clinical Neuroscience, Umeå University Hospital, Umeå, Sweden; [8]Neurodegenerative Brain Diseases Group, Department of Molecular Genetics, VIB and [9]University of Antwerp, Antwerpen, Belgium; [10]Laboratory of Neurogenetics, National Institute of Aging, NIH; [11]SDGE, National Institute of Mental Health, Bethesda, Maryland, USA; [12]Department of Genetics, University Medical Center Groningen, The Netherlands; [13]Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [14]Neuropsychiatric Institute, University of California, Los Angeles, USA; [15]These authors contributed equally to this work; [16]LHvdB and RAO are lead investigators and contributed equally to this work

ABSTRACT

We identified a SNP in the *DPP6* gene that is consistently strongly associated with susceptibility to ALS in different populations of European ancestry with an overall p-value of 5.40x10-8 in 1,767 cases and 1,916 healthy controls and with an odds ratio of 1.30 (95% CI 1.18-1.43). Our finding is the first report of a genome-wide significant association with sporadic ALS and may be a target for future functional studies.

Amyotrophic lateral sclerosis (ALS) is a severely disabling and lethal disorder caused by progressive degeneration of motor neurons in the spinal cord and brainstem. ALS affects 1-3 per 100,000 and average survival is 3 years. To date, no effective treatment is available[1].

Familial ALS accounts for up to 10% of patients with approximately 20% of families linked to mutations in *SOD1*. Mutations in *ALS2, DCTN1, VAPB* and *ANG* have been found in rare cases of familial ALS[2]. Sporadic ALS accounts for >90% of patients and is considered to be a multi-factorial disease with an estimated heritability ranging from 0.38 – 0.85[3]. Several associations with variants in genes have been reported including *ANG*[4], *VEGF*[5], *HFE*[6], *PON1*[7], and copy number variations in *SMN1/SMN2*[8]. However, replication of these findings in other populations has frequently failed. For instance, sequence variations in *ANG* are reported to be associated with ALS in Ireland and Scotland, but rarely in English and Swedish or Italian populations. Similarly, mutations in *SOD1* are found in 12-23% of families with ALS in the US, UK, Germany, Sweden and Belgium, but are rare in Portugal, Switzerland and The Netherlands (Baas, F. & Andersen, P.M. *Personal Communication* (2007)). Sporadic as well as familial ALS therefore appears to be a genetically heterogeneous disease even across European populations.

To identify novel ALS susceptibility genes we performed a genome-wide association study using Illumina 300K Beadchips (Supplementary Material). After stringent quality control, association analysis was performed on 311,946 SNPs in 461 cases and 450 healthy, age, gender and ethnically matched controls from The Netherlands. Overall call rate was 99.5%. Results for all 311,946 SNPs are available online (http://www.alscentrum.nl). No genome-wide significant association with ALS was observed after correction for multiple testing (Bonferroni).

Recently, first-stage data of a genome-wide association study of 276 ALS cases and 271 controls from the US was released. No genome-wide significant findings were observed in this study[9]. Considering the genetic heterogeneity of ALS and the fact that both studies were conducted with relatively small sample sizes, we hypothesized that the signal from truly associated SNPs might be present, although weak. We therefore decided to combine both data sets and to follow up on all SNPs with $P < 0.01$ in each study independently, and for which the allelic association was unidirectional (i.e. same direction of the allele associated). Fifteen SNPs fulfilled these criteria and were analyzed in three additional independent populations consisting of 272 cases and 336 controls from The Netherlands, 467 cases and 437 controls from Sweden and 291 cases and 420 controls from Belgium (Supplementary Material & Supplementary Table 1). Genotyping of these 15 SNPs was performed with Taqman technology (Supplementary Material). We included 100 randomly selected individuals from our Dutch GWA study sample for TaqMan genotyping of the 15 SNPs and observed a concordance rate of >99.6% between platforms. Prior to this analysis, we examined the presence of population stratification in the available GWA data from the Dutch and US sample series using Eigenstrat and failed to detect any (Supplementary Material & Supplementary Figures 1a&b)

Since the sample series in our study were derived from different populations, we calculated overall p-values and Odds Ratios (OR) using the Mantel-Haenszel method as well as the Chi-square test on allele counts for all 15 SNPs. Only one SNP, rs10260404, revealed genome-wide significance after Bonferroni correction for the 311,946 SNPs tested in the first stage. The overall p-value for rs10260404 was 5.04×10^{-8} (Corrected $P = 0.017$) with an odds ratio of 1.30 (95% CI 1.18 – 1.43) and 5.40×10^{-8} with an OR of 1.30 (95% CI 1.18 – 1.43) using the Mantel-Haenszel method[10]. The association for rs10260404 was slightly more significant under a genotypic model (Cochran-Armitage trend test) with a p-value of 3.30×10^{-8} and with an increased disease susceptibility for homozygote carriers of the risk allele (OR=1.60 with 95% CI 1.32 – 1.92) compared to heterozygotes (OR=1.20 with 95% CI 1.06– 1.41) in a dose-dependent manner.

P-values and odds ratios for each individual population are shown in Table 1. The minor allele frequency for rs10260404 was 42% for cases compared to 35% for controls. Results for all 15 SNPs analyzed in stage 2 are shown in Supplementary Table 2.

Rs10260404 maps to a 50 kb Linkage Disequilibrium block on chromosome 7q36 ($r^2>0.8$) within a gene encoding dipeptidyl peptidase 6 (*DPP6*) (Figure 1). Combining the two GWA sets revealed several SNPs within this 50 kb block that showed association with disease at $P < 0.01$. To rule out LD beyond this 50 kb block we re-examined 130 SNPs in a 900 kb region surrounding rs10260404, which did not reveal any SNP to be associated at $P < 0.01$ (Figure 1). Comparison of LD structure in this 900 kb region demonstrated

Table 1. Descriptive statistics and results for SNP rs10260404.

	n		MAF[a]		HWE[b]		P[c]	OR (95% CI)[d]
	Cases	Controls	Cases	Controls	Cases	Controls		
Stage I								
Netherlands	461	450	0.44	0.37	0.48	0.78	0.006	1.30 (1.08 - 1.56)
USA	276	271	0.42	0.34	0.28	0.06	0.003	1.45 (1.13 - 1.86)
Stage I combined*	737	721	0.43	0.36	0.39	0.19	4.30×10^{-5}	1.34 (1.08 - 1.65)
Stage II								
Netherlands	272	336	0.42	0.37	0.18	0.51	0.04	1.26 (1.01 - 1.58)
Sweden	467	439	0.4	0.34	0.26	0.23	0.006	1.31 (1.08 - 1.58)
Belgium	291	420	0.4	0.35	0.43	0.39	0.11	1.21 (0.96 - 1.51)
Stage II combined*	1030	1195	0.41	0.35	0.11	0.06	0.0002	1.26 (1.11 - 1.42)
Stage I+II combined*	1767	1916	0.42	0.35	0.33	0.38	5.40×10^{-8}	1.30 (1.18 - 1.43)

[a]Minor allele frequencies for cases and controls for each population. [b]Hardy-Weinberg Equilibrium p-values for cases and controls for each population. [c]P-values were calculated for each individual population using chi-squared test on allele counts. [d]Odds ratios (OR) were calculated for the minor allele in each population with 95% confidence interval in parenthesis. *P-values and ORs using data from multiple populations were calculated using the Mantel-Haenszel method.

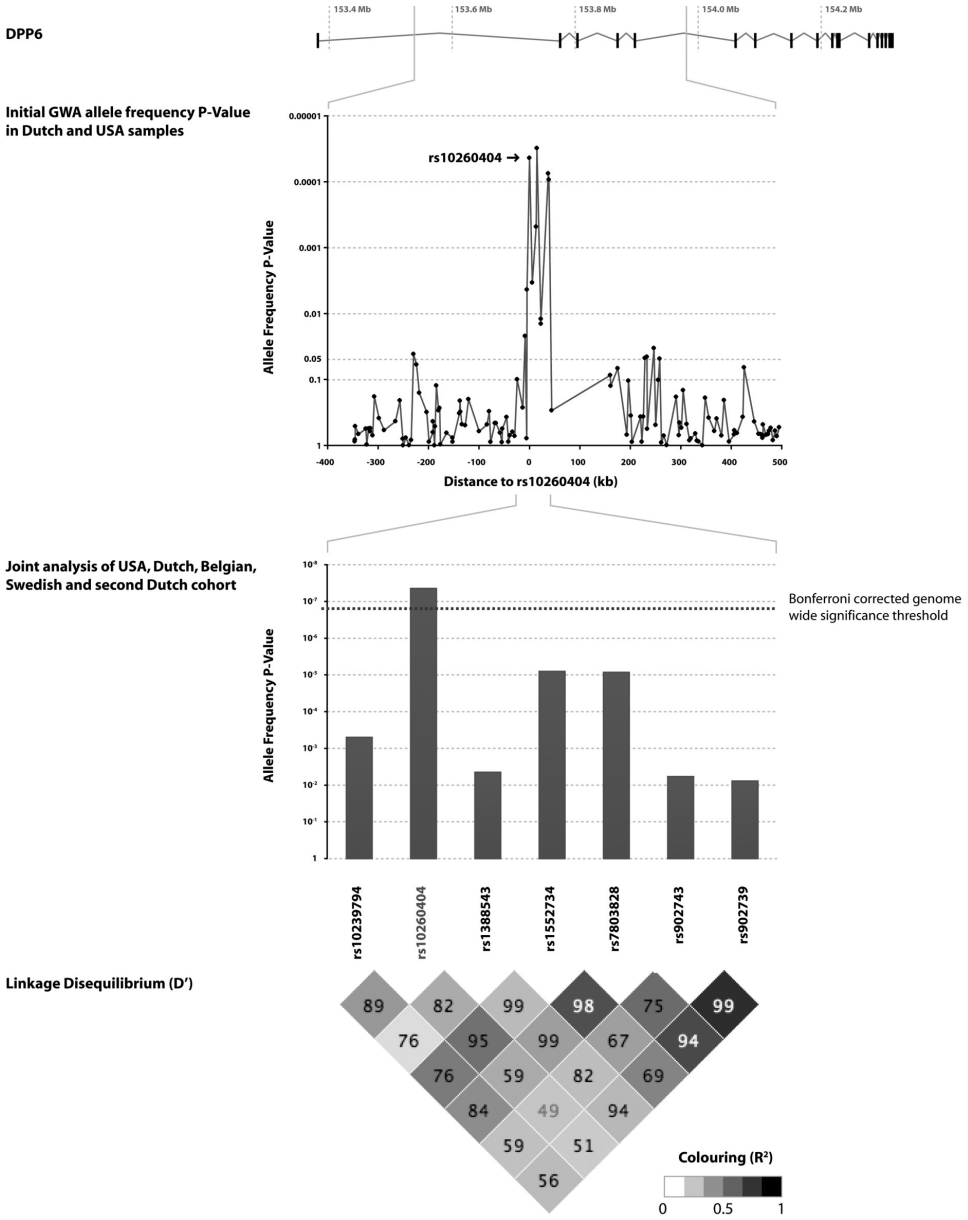


Figure 1. shows a schematic overview of DPP6. P-values from the combined analysis of the two genome-wide studies are shown for all SNPs in a 900 kb region surrounding rs10260404. rs7803828 located distal to rs10260404 demonstrated a lower p-value in the combined analysis of the GWAs, but did not fulfill the initial criteria for SNP selection (P < 0.01 in both GWAs). Subsequent analysis of 7 SNPs in the associated 50 kb locus (r^2>0.8) revealed the lowest allelic p-value for rs10260404 at P = 5.04x10^{-8}. The Bonferroni corrected genome-wide significance level was set at P = 0.05 / 311,946 = 1.6x10^{-7}.

similar haplotype structure in the Dutch, US and HapMap CEPH Sample data sets (Supplementary Figure 2a). Further examination of the associated 50 kb LD block also revealed similar LD structure in both the Dutch and US population (Supplementary Figures 2b&c). It is, therefore, unlikely that the initial finding of ALS association is due to genetic variation outside the 50 kb LD block containing rs10260404.

In order to fine-map the associated 50 kb LD block and perform haplotype analyses, we additionally genotyped all SNPs (n=6) within this block that demonstrated an association with disease at $P < 0.01$ in the combined analysis of both genome-wide studies. Genotyping of these 6 SNPs was performed with Taqman technology (Supplementary Material). Single SNP analysis of these 6 additionally genotyped SNPs did not reveal any SNP to be associated more significantly than rs10260404 (Supplementary Table 4). We then applied a recently developed multi-marker indirect association method that takes advantage of the correlation structure between SNPs in the HapMap sample (Weighted Haploype analysis (WHAP); http://whap.cs.ucla.edu/) using rs10260404 and the additional 6 flanking SNPs[11]. This imputation method yet again identified the strongest association signal for rs10260404 at $P = 6.69 \times 10^{-8}$. (Supplementary Material & Supplementary Table 5).

Subsequent haplotype analysis with Haploview using the Solid Spine of LD method to define haplotypes, revealed the strongest association signal for a haplotype containing the CC alleles of flanking SNPs rs10239794 and rs10260404 with a p-value of 3.01×10^{-9} and an allelic OR of 1.34 (95% CI 1.17 – 1.54) (Supplementary Figure 3).

Results from examining long-range LD, fine-mapping (including imputation analysis) and haplotype analysis showed the strongest signal for association hinges on the "C" allele of rs10260404, suggesting that the underlying variation for disease susceptibility is at this site. Since the entire associated 50 kb LD block containing rs10260404 is located within intron 3 of *DPP6*, without any other known gene or micro-RNA nearby, we consider this be the putative ALS-associated gene (Figure 1).

DPP6 is located on chromosome 7q36 at location 153,380,839-154,315,627 (Build 35) and consists of 26 exons and is 954 kb in size (OMIM#126141). *DPP6* (sometimes described as *DPPX*) encodes a dipeptidyl-peptidase-like protein expressed predominantly in the brain with very high levels in the amygdala, cingulated cortex, cerebellum, and parietal lobe (http://symatlas.gnf.org/SymAtlas). This peptidase regulates the biological activity of neuropeptides by converting precursors to active forms or vice versa[12]. DPP6 binds specific voltage-gated potassium channels and alters their expression and biophysical properties. Interestingly, differential DPP6 gene expression has been linked to spinal cord injury in rats[13] and *DPP6* was also identified as a nervous system-specific gene with accelerated evolutionary rate in the primate lineage[14].

In conclusion, we identified genetic variation in the *DPP6* gene that is highly associated with ALS susceptibility in a combined sample of 1,767 patients and 1,916 healthy control subjects from European descent. The identified SNP rs10260404 is located within an intron of DPP6 and no known functional variants within the gene have been identified yet. Further study will provide insight into genetic variation at this locus, its potential effect on gene function and ultimately the role it plays in disease susceptibility. Identification of a common variant within *DPP6* is an exciting first step in the genetic study of sporadic ALS, and it opens up new avenues to study the molecular basis of this devastating disease.

REFERENCES

1. Rowland, L.P. & Shneider, *N.A. N. Engl. J. Med.* 344, 1688-1700 (2001).
2. Pasinelli, P. & Brown, R.H. *Nat. Rev.* Neurosci. 7, 710-723 (2006).
3. Graham, A.J., Macdonald, A.M., & Hawkes, C.H. *J. Neurol. Neurosurg. Psychiatry* 62, 562-569 (1997).
4. Greenway, M.J. *et al. Nat. Genet.* 38, 411-413 (2006).
5. Lambrechts, D. *et al. Nat. Genet.* 34, 383-394 (2003).
6. Sutedja, N.A. *et al. Arch. Neurol.* 64, 63-67 (2007).
7. Saeed, M. *et al. Neurology* 67, 771-776 (2006).
8. Veldink, J.H. *et al. Neurology* 65, 820-825 (2005).
9. Schymick, J.C. *et al. Lancet Neurol.* 6, 322-328 (2007).
10. Skol, A.D., Scott, L.J., Abecasis, G.R., & Boehnke, M. *Nat. Genet.* 38, 209-213 (2006).
11. Zaitlen, N., Kang, H.M., Eskin, E., & Halperin, E. *Am. J. Hum.* Genet. 80, 683-691 (2007).
12. Wada, K. *et al. Mamm.* Genome 4, 234-237 (1993).
13. Tachibana, T., Noguchi, K., Ruda, M.A. *Neurosci. Lett.* 327, 133-137 (2002).
14. Dorus, S. et al. Cell 119, 1027-1040 (2004)

SUPPLEMENTARY MATERIAL

Study populations & diagnosis.
We analyzed populations from The Netherlands, USA, Belgium and Sweden. Adhering to the principles of the Declaration of Helsinki (1964), with written informed consent and approved by the local ethical committees for medical research, venous blood samples were drawn and DNA extracted according to standard procedures. All samples from Belgian and Swedish patients were screened for *SOD1* gene mutations, the Swedish samples were also screened for *ANG* mutations. No samples with mutations in these genes were included in this study. Since no *SOD1* gene mutation has ever been reported in SALS or FALS in The Netherlands, no *SOD1* screening were done in these samples.[15] All patients were diagnosed with sporadic ALS according to the 1994 El Escorial criteria.[16] Patients were included in the study when they fulfilled the criteria for probable ALS or higher. All controls had negative medical and family histories for neurodegenerative disorders. Controls were matched for age, gender and ethnicity.

Dutch population: The 461 sporadic ALS cases included in the genome wide association study were individuals referred to the University Medical Center Utrecht (UMCU), the Academic Medical Center Amsterdam (AMC) or the University Medical Center Nijmegen, St Radboud. The 450 controls included in the genome wide association study were unrelated, age- and sex-matched healthy volunteers accompanying non-ALS patients to the UMCU neurology out patient clinic and spouses of sporadic ALS patients.

272 cases and 336 controls in the second, independent Dutch population were recruited from an ongoing, prospective population based study on ALS in The Netherlands. In this study, a capture-recapture design is used to identify all prevalent and incident cases in The Netherlands. Family practitioners are then asked to recruit age- and sex-matched controls from their patient registers for each case in their practice.

Belgian population: 291 individuals with sporadic ALS were unrelated and from self reported Flemish descent for at least three generations. All patients were referred to the University Hospital Gasthuisberg, Leuven. The Belgian control group existed of 420 unrelated, healthy, Flemish individuals that were selected among married-in individuals in families with neurological diseases collected for genetic studies.

Swedish population: The Swedish cohort consisted of 467 cases and 437 controls.

Individuals with sporadic ALS were unrelated Swedish citizens who reported (northern) Swedish citizenship for at least three generations and were referred to the Umea University ALS Clinic. The Swedish control samples were spouses of the patients or unrelated healthy controls matched for age and gender.

US population: Genotype data and population characteristics for the American samples were provided digitally by Bryan J Traynor. In brief, DNA from cases and controls was obtained from the NINDS Neurogenetics Repository at the Coriell Institute for Medical Research, NJ, USA. All included individuals were unique, unrelated and from white, non-Hispanic ethnicity. 16% of sALS samples were negative for *SOD1* mutations, the remaining samples were not screened. Controls were also obtained from NINDS Neurogenetics Repository at the Coriell Institute for Medical Research and were sampled from many different regions across the US. All participants underwent a detailed medical history interview. None had a history of neurological disease.[9, 17]

Sample collection & DNA isolation

The Netherlands: Blood samples were collected in 10 ml EDTA tubes. DNA was isolated from whole blood using the autopure DNA isolation protocol from Qiagen (Qiagen, Valencia, USA).

USA: DNA from cases and controls was obtained from the NINDS Neurogenetics Repository at the Coriell Institute for Medical Research, NJ, USA. DNA was extracted from Epstein-Barr virus immortalized lymphocyte cell lines using a salting out procedure.
Belgium: Genomic DNA was extracted from peripheral lymphocytes using standard procedures, on a Chemagen Magnetic Separation Module 1 platform (Chemagen AG, Baesweiler, Germany).

Sweden: Genomic DNA was extracted from white blood cells from peripheral blood using the QIAamp Blood Kit (Qiagen).

Genotyping methods

For the Dutch genome wide association study we genotyped 461 sALS cases and 450 healthy controls derived from the Dutch population. Genotyping experiments were performed at the University Medical Centre Utrecht using the Illumina Infinium II HumanHap300 SNP chips (Illumina, San Diego, CA).

The sALS samples from the US were assayed using the Illumina Infinium II HumanHap 550 SNP chips. 227 controls from the US were genotyped using the Illumina Infinium II HumanHap300 SNP chips and typed additionally using the Illumina 240S chip. A further 48 controls were assayed using the Illumina Infinium II HumanHap 550 SNP chip.
All experiments were carried out according to the manufacturer's protocol. In short, 750 ng of DNA per sample was whole-genome amplified, fragmented, precipitated and resuspended in the appropriate hybridization buffer. Subsequently denatured samples were hybridized on Illumina BeadChips at 48°C for a minimum of 16 hours. After hybridization, the beadchips were processed for the single base extension reaction and stained. Chips were then imaged using the Illumina Bead Array Reader. For each sample, normalized bead intensity data was loaded into Illumina Beadstudio 2.0 and converted into genotypes. Genotypes were called using the auto-calling algorithm in Illumina Beadstudio 2.0.

The SNPs, which were selected for replication in additional populations, were genotyped using Taqman allelic discrimination assays. PCR was carried out with mixes consisting of 10 ng of genomic DNA, 1 x Taqman master mix (Applied Biosystems), 1x assay mix (Applied Biosystems, Foster City, USA) and ddH2O in a 5 μl reaction volume in 384-well plates (Applied Biosystems). PCR conditions were as follows: denaturation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 92 °C for 15 seconds and

annealing and extension at 60°C for 1 minute.

Allelic PCR products were analyzed on the ABI Prism 7900HT Sequence Detection System using SDS 2.3 software (Applied Biosystems). Primer and probe sequences are available upon request.

We genotyped 100 random individuals from the Dutch genome wide association study for all 15 selected SNPs using the Taqman allelic discrimination assays to ensure both SNP genotyping platforms generated the same genotypes for each individual (concordance rate = 99.6%).

Quality control

Extensive quality control was done on the data from Dutch genome-wide study before performing statistical analysis. In the genome-wide study we genotyped 477 unique samples from Dutch sporadic ALS cases and 472 unique Dutch controls. We excluded 34 individuals (12 cases and 22 controls) for poor quality genotyping (call rate < 95%). Two pairs of cases (n=4) were excluded due to observed family relationships (> 200.000 concordant SNPs). Two samples were genotyped twice yielding concordance of over 99.9% for each sample. The average call rate across all samples was 99.5%. The call rate was ≥ 99% for 298,807 SNPs and ≥ 95% for 315,293 SNPs. HWE was calculated in controls for each SNP and ≤ 0.05 for 14,498 SNPs and ≤ 0.01 for 3,207 SNPs. The average minor allele frequency was 26.1%. 396 SNPs had a minor allele frequency < 0.01. After pruning SNPs according to frequency (MAF < 0.01) and genotyping (missingness ≥ 0.05, HWE ≤ 0.01) 311,946 SNPs remained. A total of 284,182,806 unique genotype calls were made.

For the US genome-wide study quality control was performed as previously described[4]. In short, 276 unique samples from sporadic ALS cases and 275 unique controls were genotyped. A total of 3 controls were excluded for poor quality genotyping (call rate < 95%) and 1 individual was excluded for having an African American background. One sample was genotyped twice yielding concordance of over 99.9%. The average call rate across all samples was 99.6%. The call rate was ≥ 99% for 514,088 SNPs and ≥ 95% for 549,062 SNPs. HWE was calculated for each SNP and was ≤ 0.05 for 23,657 SNPs and ≤ 0.01 for 5,911 SNPs. The average minor allele frequency was 23.7%. A total of 302,655,011 unique genotype calls were made. Only SNPs on the Illumina 300K chips with a call rate > 95% and MAF >0.01 were included for statistical analysis.

Statistical analysis

Association analysis: Illumina Beadstudio 2.0 was used to generate genotype final report files, which we converted into pedigree and map files using the software program Perl (http://www.perl.com). Subsequently, pedigree files were loaded into PLINK 0.99s to perform association analysis using the chi-squared test on allele counts. To calculate p-values over genotype data derived from different populations we used the Maentel-Haenszel method using PLINK 0.99s.[18] P-values for the selected 15 SNPs were also calculated under different models with PLINK 0.99s.

Considering the Dutch genome-wide association study was performed using the Illumina Hap300 chip with roughly 317,000 SNPs (compared to the US GWA which included >500,000 SNPs, only the Hap300 SNPs were included for the analyses. We further excluded SNPs from analysis if: call rate <95% in either case or control samples in the US or Dutch data sets, minor allele frequency <1% in either US or Dutch sample series or for deviation from Hardy-Weinberg equilibrium (P < 0.0001 in control samples). A total of 311,946 SNPs were included for analysis. This number was also used when correcting for multiple testing using the Bonferroni method. Odds ratios with 95% CI

were calculated for the minor allele of each SNP.
Power calculations: Power calculation for the genome-wide studies were performed using Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/) based on sample size, the average observed minor allele frequency under the assumption of a multiplicative model and a prevalence of 2 per 100,000.[19]
Using these parameters the Dutch genome scan was 80% powered to detect an allelic association with $P < 0.01$ and an odds ratio of 1.44. The US genome scan was 80% powered to detect an allelic association with $P < 0.01$ and an odds ratio of 1.57. Power was calculated for stage I using the p-value of 0.01, since this was the cut-off value set for SNP selection. In Stage II of this study 15 SNPs were selected for further analysis. We calculated the power for replication for each of these SNPs in Stage II based on the observed MAF in Stage I and odds ratios at a p-value of 0.05 and the Bonferroni corrected value of $P = 0.05 / 15 = 0.0033$. The size of the overall replication sample provided at least 80% power to detect for each of the 15 SNPs at a p-value of 0.05 and the corrected level of 0.0033 (Results not shown). Supplementary Table 3 shows the overall power for the study at P = 0.01, 1.0×10^{-5} and 1.6×10^{-7} (which corresponds to genome-wide significance P = 0.05 / 311,946).

Haplotype analysis: Haploview v3.32 was used for assessing linkage disequilibrium patterns and haplotype association statistics.[20] Haplotypes were defined using the solid spine of LD setting in Haploview. We included all haplotypes with a frequency>1%. P-values were calculated using a Chi-square test on alleles. Haplotypes were estimated using an accelerated EM algorithm, creating highly accurate estimations of the population frequencies of the phased haplotypes based on the maximum likelihood ratio as determined from the unphased input.

Results from population stratification analysis
We analyzed the data from both genome-wide studies for evidence of population stratification between cases and controls using Eigenstrat (Supplementary Figure 1). P-values generated by Eigenstrat were essentially identical to the initial analysis and the ranking of SNPs remained the same. Furthermore, we calculated a genomic control population inflation factor between cases and controls of 1.00. Both methods revealed no evidence for stratification between cases and controls in the genome-wide association studies.

We calculated a genomic control population inflation factor between the Dutch and US population of 1.00, indicating no evidence for strong population stratification between the two populations. Furthermore, the case-control ratio in both studies was identical, hereby further reducing possible influence due to population stratification (461 : 450 = 1.02, cases versus controls for The Netherlands compared to 276 / 271 = 1.02, cases versus controls for the US).

Even though we did not observe any evidence for population stratification, the Mantel-Haenszel method was used to calculate p-values when combining data from different populations, hereby taking into account that our overall study sample was comprised of subjects from The Netherlands, United States, Belgium and Sweden and could represent different strata.

Results: Additional genotyping and Haplotype analysis
Considering only one SNP in the associated locus fulfilled our criteria for follow-up ($P < 0.01$ in both GWAS), we hypothesized that other SNPs or haplotypes could be associated at a more significant level in the overall sample (including Stage II populations), while not necessarily most significant in the discovery set (Stage I). We therefore selected

six additional tagging SNPs within the associated 50 kb locus that demonstrated an association with disease at $P < 0.01$ in the combined analysis of both genome-wide studies and subsequently genotyped these SNPs in all populations. The 6 additional SNPs were rs10239794, rs1388543, rs1552734, rs7803828, rs902743 and rs902739.

Individual testing of these SNPs from this region showed that none was more significantly associated than the previously identified rs10260404. However, all additionally typed SNPs consistently demonstrated p-values < 0.05. Results are shown in Supplementary Table 4. Using the solid spine setting in Haploview, 2 haplotype blocks were identified. Subsequent association analysis demonstrated a p-value of 3.01×10^{-9} for the CC-haplotype of rs10260404 and rs10239794 (Supplementary Figure 3).

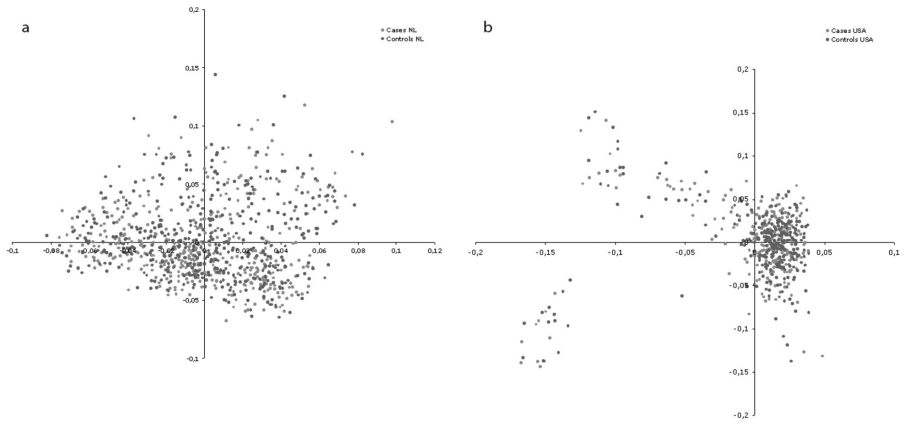
Weighted Haplotype analysis (WHAP)

WHAP is a statistical test for case/control association studies. It takes advantage of the known correlation structure between SNPs in the HapMap sample (http://www.hapmap.org) in order to improve power over traditional methods.[21]

The full details of the rationale and the methodology are described in Zaitlen et al.[11] The WHAP tool is available online at http://whap.ucla.edu/. The analysis was performed using genotype data from all included subjects from all populations for the following SNPs: rs10239794, rs10260404, rs1388543, rs1552734, rs7803828, rs902743 and rs902739. Quality control parameters were set at 0.001 for the Hardy-Weinberg Equilibrium p-value and 10% for missing genotype or missing individual frequencies. HapMap correlations were taken from the European-American (CEPH) HapMap sample. WHAP analysis yielded the most significant association with the collected marker rs10260404 at $P = 6.69 \times 10^{-8}$ (Supplementary Table 5). The most significant result from the WHAP analysis is the same SNP identified by the initial scan and fine mapping, suggesting that the true association signal represented with these seven SNPs and the imputed haplotype structure is located at very close proximity to this marker or rs10260404 itself.

REFERENCES

15. Baas,F. & Andersen,P.M. *Personal Communication* (2007).
16. Brooks,B.R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J. Neurol. Sci.* 124 Suppl, 96-107 (1994).
17. Fung,H.C. et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 5, 911-916 (2006).
18. Purcell,S. PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 81, 559-755 (2007).
19. Purcell,S., Cherny,S.S., & Sham,P.C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 19, 149-150 (2003).
20. Barrett,J.C., Fry,B., Maller,J., & Daly,M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263-265 (2005).
21. The International HapMap Consortium. The International HapMap Project. *Nature* 426, 789-796 (2003).



Supplementary Figure 1. *(a) Eigenstrat analysis between Dutch cases and controls shows no evidence for population stratification. Cases are shown in light grey and controls in dark grey. (b) Eigenstrat analysis between US cases and controls shows evidence for structure, but no evidence for population stratification. Cases are shown in light grey and controls in dark grey.*

Block 1

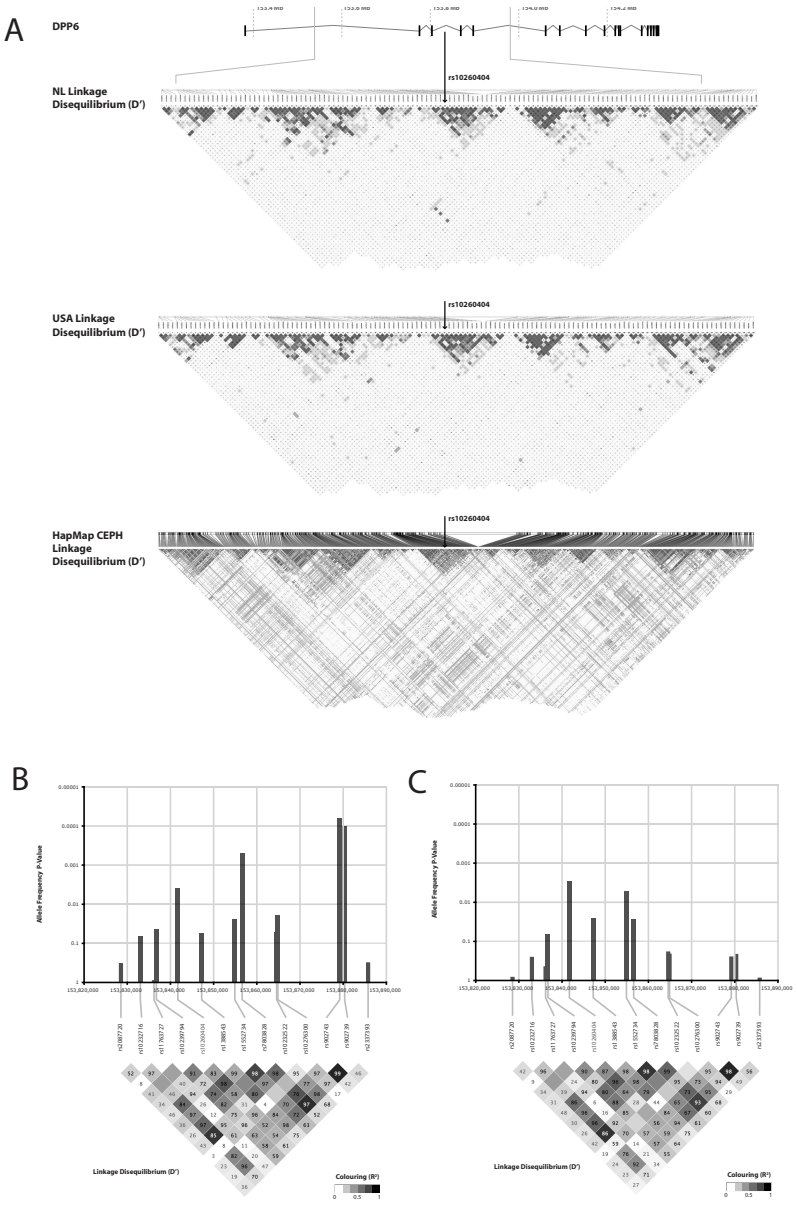
rs10239794 rs10260404	Case Freq	Control Freq	P value
TT	0.385	0.421	0.0018
CC	0.401	0.334	**3.01 x 10^{-9}**
CT	0.199	0.226	0.0056
TC	0.015	0.02	0.1514

D' 0.69

Block 2

rs1388543 rs1552734 rs7803828 rs902743 rs902739	Case Freq	Control Freq	P value
TTTCT	0.384	0.421	0.0013
TCGTC	0.202	0.18	0.0221
CCGTC	0.197	0.169	0.0021
TCTCT	0.092	0.089	0.7085
TTTTC	0.059	0.074	0.0087
TCGCC	0.032	0.03	0.6396
CCGCC	0.015	0.017	0.5209

Supplementary Figure 3. *Using the Solid Spine of LD method in Haploview two haplotype blocks were defined. The CC-haplotype composed of rs10239794 and rs10260404 demonstrated the most significant result.*



Supplementary Figure 2. *(a) shows a similar haplotype structure in the Dutch (NL), US and in the HapMap CEPH sample. Rs10260404 is located in a block with relatively low LD, therefore it seems unlikely that the initially observed association signal was derived from genetic variation outside of the 50 kb block of LD (r^2>0.8) in which rs102060404 is located. (b) Top part of the figure shows P-values for SNPs in the 50 kb LD block surrounding rs10260404 in the Dutch population. The bottom part shows LD structure. (c) Top part of the figure shows P-values for SNPs in the 50 kb LD block surrounding rs10260404 in the US population. The bottom part shows LD structure.*

Supplementary Table 1. shows baseline characteristics for the studied population

Study Populations	Total	Male (%)	Spinal Onset (%)	Age at Onset (yrs)*	Survival (months)**
The Netherlands, GWA (Stage I)					
ALS	461	59	69	59 (20-86)	33 (5-147)
Healthy Controls	450	59	-	60 (22-87)	-
USA, GWA (Stage I)					
ALS	276	63	78	55 (26-87)	-
Healthy Controls	271	48	-	62 (25-94)	-
The Netherlands, 2nd population (Stage II)					
ALS	272	57	71	58 (16-83)	43 (4-196)
Healthy Controls	336	57	-	59 (29-95)	-
Belgium (stage II)					
ALS	291	59	73	59 (18-86)	39 (5-177)
Healthy Controls	420	58	-	51 (18-92)	-
Sweden (Stage II)					
ALS	467	57	66	60 (20-89)	31 (9-108)
Healthy Controls	439	52	-	62 (25-94)	-
Total					
ALS	1767	58	71	59 (16-89)	36 (4-196)
Healthy Controls	1916	57	-	58 (18-95)	-

*Age at onset is shown in years with range in parenthesis. **Survival is shown in months with range shown in parenthesis. No survival data is available for the US cohort.

Supplementary Table 2. shows P-values for all 15 SNPs with P < 0.01 in Stage I.

SNP	Chr	Gene	GWA NL *P*	GWA USA *P*	NL Stage II *P*	Be Stage II *P*	Swe Stage II *P*	Combined Stage II *P*	MAF		Overall *P*	OR (9% CI)
									Cases	Controls		
rs10260404	7	*DPP6*	0.006	0.003	0.04	0.11	0.006	0.004	0.42	0.35	$5.04x10^{-8}$	1.30 (1.18 - 1.43)
rs3825776	15	*LIPC*	0.008	0.003	0.005	0.71	0.21	0.009	0.35	0.29	$8.75x10^{-6}$	1.34 (1.20-1.46)
rs7580332	2	No Gene	0.005	0.002	0.36	0.56	0.08	0.08	0.40	0.45	$8.78x10^{-6}$	0.82 (0.74-0.92)
rs973807	8	*NSMAF*	0.0002	0.006	0.81	0.74	0.50	0.68	0.30	0.35	0.0006	0.82 (0.73-0.92)
rs1061947	17	*COL1A1*	0.005	0.002	0.49	0.82	0.93	0.74	0.18	0.15	0.002	1.23 (1.08-1.40)
rs9409314	9	No Gene	0.001	0.010	0.70	0.85	0.44	0.86	0.38	0.34	0.003	1.16 (1.05-1.29)
rs9380343	6	No Gene	0.003	0.001	0.85	0.84	0.77	0.88	0.03	0.05	0.003	0.69 (0.54-0.88)
rs10438933	18	No Gene	0.003	0.003	0.20	0.35	0.21	0.93	0.14	0.12	0.004	1.24 (1.07-1.43)
rs5924655	23	*PASD1*	0.009	0.003	0.57	0.68	0.49	0.80	0.25	0.28	0.006	0.81 (0.66-0.93)
rs1574549	7	*CALN1*	0.004	0.001	0.44	0.93	0.18	0.77	0.32	0.29	0.008	1.15 (1.04-1.28)
rs1493282	17	*ASPA*	0.007	0.002	1.00	0.50	0.79	0.95	0.26	0.23	0.007	1.17 (1.04-1.31)
rs12861395	23	*PASD1*	0.007	0.003	1.00	0.31	0.41	0.28	0.25	0.28	0.03	0.85 (0.72-0.97)
rs895459	2	*BARD1*	0.010	0.003	0.11	0.67	0.44	0.68	0.37	0.39	0.07	0.91 (0.81-1.01)

rs1123319	23	*PASD1*	0.004	0.001	1.00	0.31	0.41	0.45	0.22	0.25	0.08	0,83 (0,69-1,02)
rs11127401	2	No Gene	0.010	0.008	0.03	0.06	0.61	0.06	0.32	0.33	0.32	0.88 (0.82-1.06)

NL= The Netherlands, Be= Belgium, Swe= Sweden. Overall MAF= minor allele frequency calculated over all 5 populations. OR= odds ratio, shown for each SNPs minor allele with 95% confidence interval shown in parentheses.

Supplementary Table 3. Overall Power

MAF	OR	Power (%)		
		P = 0.01	P = 1.0 x10^{-5}	P = 1.6x10^{-7}
0.10	1.30	84	20	5
	1.50	100	90	65
	1.70	100	99	99
0.35	1.30	100	85	58
	1.50	100	100	100
	1.70	100	100	100

Power is shown calculated over Stage I&II, total of 1,767 cases and 1,916 controls, at different minor allele frequencies (MAF) and different odds ratios (OR) at 3 separate p-values. A p-value of 1.6 x10^{-7} corresponds to genome-wide significance after Bonferroni correction.

Supplementary Table 4. Results for additionally genotyped SNPs

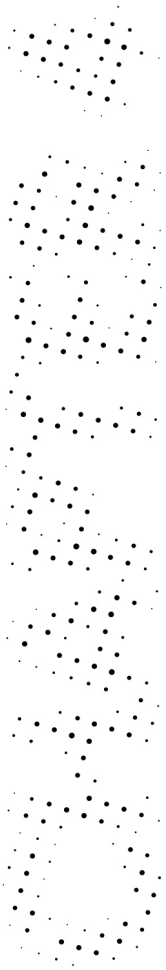
SNP	Stage I *P*			Stage II *P*				Overall *P*
	NL	USA	Meta-analysis *P*	NL	Swe	Be	Meta-analysis *P*	
rs10239794	0.045	0.039	0.004	0.29	0.06	0.54	0.06	6.10x10^{-4}
rs1388543	0.053	0.02	0.003	0.24	0.63	0.3	0.43	0.009
rs1552734	0.034	0.003	5. 0x10^{-4}	0.15	0.11	0.11	0.005	7.44x10^{-6}
rs7803828	5.53x10-4	0.019	3.0x10^{-5}	0.31	0.23	0.42	0.31	3.08x10^{-5}
rs902743	5.90x10-5	0.196	7.5x10^{-5}	0.37	0.4	0.84	0.99	0.01
rs902739	1.01x10-4	0.173	9.0x10^{-5}	0.63	0.52	0.8	0.87	0.01

NL= The Netherlands, Swe= Sweden, Be= Belgium. P-values were calculated using the Chi-square test on allele counts. Overall P-values were calculated using the Mantel-Haenszel Method.

Supplementary Table 5. Results from WHAP analysis.

SNP	Chr	Chr pos	MAF	r^2	P	Tag set
rs10260404	**7**	**153,841,730**	**0.38**	**1.00**	**6.69×10^{-8}**	**collected marker**
rs10264387	7	153,851,755	0.50	0.94	3.75×10^{-7}	rs10260404,rs1552734,rs902743
rs12668377	7	153,840,791	0.41	0.97	1.03×10^{-6}	rs10260404,rs1388543
rs10247061	7	153,837,975	0.43	0.85	5.69×10^{-6}	rs10239794,rs10260404,rs902739
rs11975187	7	153,863,909	0.39	1.00	6.84×10^{-6}	rs10260404,rs7803828,rs902743
rs1907616	7	153,859,314	0.40	0.97	7.15×10^{-6}	rs10260404,rs7803828,rs902743
rs6956780	7	153,850,432	0.21	0.68	7.43×10^{-6}	rs10260404,rs1388543,rs1552734
rs7803828	**7**	**153,856,587**	**0.43**	**1.00**	**1.19×10^{-5}**	**collected marker**
rs1552734	**7**	**153,854,973**	**0.46**	**1.00**	**1.51×10^{-5}**	**collected marker**
rs923711	7	153,852,675	0.45	0.97	1.56×10^{-5}	rs1388543,rs1552734,rs902743
rs10952491	7	153,848,888	0.50	0.86	3.43×10^{-5}	rs10260404,rs7803828,rs902743
rs6964455	7	153,839,596	0.48	0.83	3.63×10^{-5}	rs10239794,rs10260404,rs1552734
rs13246178	7	153,848,111	0.18	0.85	3.90×10^{-5}	rs10260404,rs1388543,rs1552734
rs12538549	7	153,840,163	0.18	0.75	4.86×10^{-5}	rs10260404,rs1388543,rs1552734
rs7780519	7	153,848,685	0.18	0.85	5.18×10^{-5}	rs10260404,rs1388543,rs1552734
rs13232525	7	153,839,757	0.18	0.71	6.16×10^{-5}	rs10260404,rs1388543,rs1552734
rs11767475	7	153,839,973	0.18	0.75	6.58×10^{-5}	rs10260404,rs1388543,rs1552734
rs6464429	7	153,844,747	0.43	0.90	9.61×10^{-5}	rs10239794,rs10260404,rs902743
rs12534820	7	153,834,965	0.16	0.70	2.02×10^{-4}	rs10239794,rs10260404,rs1388543
rs4725547	7	153,834,387	0.43	0.86	2.20×10^{-4}	rs10239794,rs1388543,rs902743
rs10267199	7	153,834,908	0.40	0.89	2.42×10^{-4}	rs10239794,rs7803828
rs1388540	7	153,853,562	0.39	0.76	2.55×10^{-4}	rs10260404,rs7803828,rs902743
rs10276300	7	153,864,766	0.10	0.75	2.58×10^{-4}	rs10260404,rs1388543,rs902743
rs10262182	7	153,833,200	0.45	0.87	4.45×10^{-4}	rs10239794,rs902743,rs902739
rs10952485	7	153,829,535	0.43	0.93	6.3×10^{-4}	rs10239794,rs10260404,rs1552734
rs10239794	**7**	**153,836,826**	**0.42**	**1.00**	**6.89×10^{-4}**	**collected marker**
rs11976788	7	153,829,985	0.49	0.97	7.59×10^{-4}	rs10239794,rs902743,rs902739
rs6955044	7	153,832,247	0.50	0.94	1.10×10^{-3}	rs10239794,rs902743,rs902739
rs9690048	7	153,852,822	0.08	0.90	1.31×10^{-3}	rs10260404,rs1388543,rs902743
rs7809068	7	153,859,565	0.08	0.90	1.31×10^{-3}	rs10260404,rs1388543,rs902743
rs11971700	7	153,864,956	0.08	0.90	1.31×10^{-3}	rs10260404,rs1388543,rs902743
rs13228346	7	153,866,129	0.08	0.90	1.31×10^{-3}	rs10260404,rs1388543,rs902743
rs12703363	7	153,846,598	0.08	0.90	1.31×10^{-3}	rs10260404,rs1388543,rs902743
rs13226031	7	153,832,786	0.26	0.76	1.58×10^{-3}	rs10239794,rs10260404,rs1388543
rs10260955	7	153,838,166	0.34	0.82	2.41×10^{-3}	rs10260404,rs1388543,rs7803828
rs10274497	7	153,869,441	0.20	0.87	2.73×10^{-3}	rs10239794,rs1388543,rs7803828
rs923710	7	153,852,787	0.19	0.83	2.96×10^{-3}	rs10239794,rs1388543,rs7803828
rs10231561	7	153,837,944	0.18	1.00	4.02×10^{-3}	rs10239794,rs1388543
rs10464419	7	153,856,052	0.30	0.77	4.15×10^{-3}	rs10239794,rs1552734,rs902743
rs6969351	7	153,877,668	0.46	0.97	4.27×10^{-3}	rs10260404,rs902743,rs902739
rs12533032	7	153,834,856	0.28	0.74	4.60×10^{-3}	rs10239794,rs10260404,rs1388543
rs11766937	7	153,836,441	0.28	0.74	4.6×10^{-3}	rs10239794,rs10260404,rs1388543
rs1120724	7	153,837,511	0.08	0.63	4.75×10^{-3}	rs10239794,rs1552734,rs902739
rs1388543	**7**	**153,847,278**	**0.20**	**1.00**	**6.29×10^{-3}**	**collected marker**
rs12703360	7	153,831,979	0.29	0.74	6.55×10^{-3}	rs10239794,rs10260404,rs1388543
rs6976924	7	153,850,457	0.13	1.00	6.85×10^{-3}	rs10239794,rs10260404,rs1552734
rs902739	**7**	**153,880,344**	**0.50**	**1.00**	**8.35×10^{-3}**	**collected marker**
rs902743	**7**	**153,879,200**	**0.45**	**1.00**	**8.74×10^{-3}**	**collected marker**

Results from WHAP analysis showing the top 50 SNP locations; the collected markers are shown in bold while the others are all imputed based on haplotype correlations present in the European-American CEPH HapMap sample.



GENOME-WIDE ASSOCIATION STUDY IDENTIFIES 19P13.3 *(UNC13A)* AND 9P21.2 AS SUSCEPTIBILITY LOCI FOR SPORADIC AMYOTROPHIC LATERAL SCLEROSIS

Nature Genetics (2009)

Michael A van Es[1,27], Jan H Veldink[1,27], Christiaan GJ Saris[1], Hylke M Blauw[1], Paul WJ van Vught[1], Anna Birve[2], Robin Lemmens[3], Helenius J Schelhaas[4], Ewout JN Groen[1], Mark HB Huisman[1], Anneke J van der Kooi[5], Marianne de Visser[5], Caroline Dahlberg[2], Karol Estrada[6], Fernando Rivadeneira[6], Albert Hofman[11], Machiel J Zwarts[4], Dan Rujescu[7], Eric Strengman[8], Ina Giegling[7], Pierandrea Muglia[9], Barbara Tomik[10], Agnieszka Slowik[10], Andre G Uitterlinden[6,11], Corinna Hendrich[12], Stefan Waibel[12], Thomas Meyer[13], Albert C Ludolph[12], Jonathan D Glass[14], Shaun Purcell[15], Sven Cichon[16], Markus M Nöthen[16], H-Erich Wichmann[17], Stefan Schreiber[18], Sita HHM Vermeulen[19], Lambertus A Kiemeney[20], John HJ Wokke[1], Simon Cronin[21], Russell L McLaughlin[21], Orla Hardiman[21], Katsumi Fumoto[22], R Jeroen Pasterkamp[22], Vincent Meininger[23], Judith Melki[24], P Nigel Leigh[25], Christopher E Shaw[25], John E Landers[26], Ammar Al-Chalabi[25], Robert H Brown Jr[26], Wim Robberecht[3], Peter M Andersen[2], Roel A Ophoff[8,27], Leonard H van den Berg[1,27]

[1]Department of Neurology, Rudolf Magnus Institute, UMC Utrecht, The Netherlands; [2]Institute of Clinical Neuroscience, Umeå University Hospital, Umeå, Sweden; [3]Department of Neurology, University Hospital Leuven, University of Leuven, Belgium; [4]Department of Neurology, RUNMC, Nijmegen, The Netherlands; [5]Department of Neurology, Amsterdam Medical Center, Amsterdam, The Netherlands; [6]Department of Internal Medicine, Erasmus MC Rotterdam, The Netherlands; [7]Department of

Psychiatry, Ludwig-Maximilians-University, Munich, Germany; [8]Department of Medical Genetics, UMC Utrecht, The Netherlands; [9]Medical Genetics, GlaxoSmithKline R&D, Verona, Italy; [10]Department of Neurology, Jagiellonian University, Krakow, Poland; [11]Department of Epidemiology, Erasmus MC Rotterdam, The Netherlands; [12]Department of Neurology, University of Ulm, Ulm, Germany; [13]Department of Neurology, Charité University Hospital, Berlin, Germany; [14]Department of Neurology, Emory University, Atlanta, USA; [15]Center for Human Genetics Research, Massachusetts General Hospital, Boston, USA; [16]Institute of Human Genetics, University of Bonn, Bonn, Germany; [17]Institute of Epidemiology, Helmholtz Center Munich, Germany; [18]Institute for Clinical Molecular Biology, University of Kiel, Kiel, Germany; [19]Dept. of Human Genetics, RUNMC, Nijmegen, The Netherlands; [20]Department of Urology, RUNMC, Nijmegen, The Netherlands; [21]Department of Neurology, Beaumont Hospital, Dublin, Ireland; [22]Dept. Neuroscience & Pharmacology, RMI, UMC Utrecht, The Netherlands; [23]Department of Neurology, Pitié Salpêtrière Hospital, Paris, France; [24]Department of Human Genetics, Hadassah University Hospital, Jerusalem, Israel; [25]Department of Clinical Neuroscience, King's College, IoP, London, UK; [26]Department of Neurology, Massachusetts General Hospital, Charlestown, USA; [27]These authors contributed equally to this work.

ABSTRACT

We performed a genome-wide association study in sporadic ALS (2,323 patients and 9,013 controls) and evaluated all SNPs with $P < 1.0 \times 10^{-4}$ in a second, independent cohort of 2,532 patients and 5,940 controls. Analysis of the genome-wide data revealed genome-wide significance for one SNP, rs12608932 with $P = 1.30 \times 10^{-9}$. This SNP demonstrated robust replication in the second cohort ($P = 1.86 \times 10^{-6}$) and a combined analysis over the two stages yielded a $P = 2.53 \times 10^{-14}$. The rs12608932 SNP is located at 19p13.3 and maps to a haplotype block within the boundaries of *UNC13A*, which regulates the release of neurotransmitters, such as glutamate, at neuromuscular synapses. Follow-up of additional SNPs demonstrated genome-wide significance for two further SNPs (rs2814707 with $P = 7.45 \times 10^{-9}$ and rs3849942 with $P = 1.01 \times 10^{-8}$) in the combined analysis of both stages. These SNPs are located at chromosome 9p21.2 in a linkage region for familial ALS with fronto-temporal dementia found previously in several large pedigrees.

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a neurodegenerative disorder characterized by progressive wasting and weakness of limb, bulbar, and respiratory muscles. The disease is caused by loss of motor neurons in the spinal cord, brainstem, and motor cortex and can occur at any time in adulthood, with a median age of onset in the mid-fifties. The incidence of sporadic ALS is 1.5–2.8 per 100,000 population per year with a male predominance (ratio about 1.6:1). About half of patients die within 3 years of symptom onset, because of respiratory failure.[1] Currently the only medication to slow the disease is riluzole, which delays progression by 3 months.[2] The genetic contribution to sporadic ALS remains largely unknown, despite the considerable estimated heritability ranging from 0.35 to 0.85.[3]

Candidate gene studies have identified potential risk factors, but replication of these associations has proven to be difficult. Similarly, several genome-wide association (GWA) studies have been performed in sporadic ALS, but have not identified irrefutable associations. *ITPR2*, *FGGY* and *DPP6* have been proposed as candidate susceptibility

loci. However, these association findings were not consistently replicated in other GWA studies.[4-10] The GWA studies in ALS have been relatively small and as a result of limited power may have resulted in potentially spurious results.

In order to identify truly associated genetic risk factors for ALS, we performed a two-stage GWA study encompassing 19,838 subjects. An initial genome-wide screen with 2,323 patients and 9,013 controls was followed by a replication experiment of all SNPs with $P < 1.0 \times 10^{-4}$ in an independent cohort of 2,532 patients and 5,940 controls. Genotyping experiments were performed using Illumina BeadChips and KASPar genotyping technology. A detailed description of the samples series for the genome-wide screen and follow-up stage is provided in Table 1 and Supplementary Information.

This study design provided equal power to detect disease associations in the replication and genome-wide phases. Power calculations were performed using Genetic Power Calculator (http://pngu.mgh.harvard.edu/~purcell/gpc/). The GWA and replication study had 90.6% and 92.9% power respectively to detect a disease association with OR = 1.30 at $P = 5.0 \times 10^{-8}$ assuming a disease prevalence of 8.75 per 100,000 population, a multiplicative model and a risk allele frequency of 20% (Supplementary Figure 1).

Considering the power of this study to detect disease associations for SNPs with OR < 1.20 at $P = 5.0 \times 10^{-8}$ was limited (Supplementary Figure 1), we additionally selected all SNPs with $P < 1.0 \times 10^{-4}$ for follow-up. This approach increases the sensitivity of the study and allowed us to select more true positives that would not be detected at higher significance thresholds, but comes at the expense of taking forward more false positives.

After applying quality control and stratification analysis to the genome-wide data set, genotypic data was available for 292,768 SNPs in a total of 2,323 patients and 9,013 controls (Supplementary Table 1). Association testing was performed with the statistical analysis program PLINK v1.05[11] using logistic regression with adjustment for gender, nationality (dummy coded) and ancestry. Gender was included as covariate accounting for the male preponderance in ALS and correcting for the variable distribution of gender in the different cohorts. For ancestry, we used the first 2 dimensions of the multidimensional scaling on the matrix of identity-by-state sharing of all pairs of individuals as covariates. A quantile-quantile plot for the genome-wide phase, in which the observed –log10 p-values are plotted against the expectation under the null hypothesis, shows an excess of significant associations ($\lambda GC=1.05$) (Supplementary Figure 2), with one SNP exceeding the genome-wide threshold of $P = 5.0 \times 10^{-8}$ (rs12608932 with $P = 1.30 \times 10^{-9}$).[12] In the genome-wide phase we observed a total of 58 SNPs with $P < 1.0 \times 10^{-4}$, where 29 SNPs were expected. This enrichment suggested the presence of potentially multiple true positives amongst these SNPs. We therefore selected these 58 SNPs for further analysis in the replication phase.

The previously associated SNP in *DPP6* (rs10260404) met the criteria for follow-up with $P = 9.23 \times 10^{-5}$. Although the reportedly associated SNPs in *ITPR2* and *FGGY* did not yield $P < 1.0 \times 10^{-4}$, they were selected for follow-up genotyping. Therefore, a total of 60 SNPs was analyzed in the second stage (Supplementary Table 2).

After applying standard quality control procedures to the replication phase, genotype data for 2,532 patients and 5,940 controls was available for analysis (Supplementary Table 3). Confounding due to population stratification was a serious concern given the multiple nationalities of the populations across patient and control cohorts (Table 1). We therefore corrected for population stratification in the analysis of the genome-wide data as well as in the replication cohort. In the genome-wide phase

Table 1. Study populations and genotyping platforms

Genome-wide association study:

Country	*n* Patients	*n* Controls	Genotyping platform	Genotyped by	Source
The Netherlands	461	450	Illumina 317K	UMC Utrecht	UMC Utrecht; previously published[4, 5]
The Netherlands	585	603	Illumina 370K	deCODE	UMC Utrecht; genotyped
The Netherlands	-	677	Illumina 550K	UCLA	UMC Utrecht; in silico[21]
The Netherlands	-	5,974	Illumina 550K	Erasmus MC	RS-I cohort, The Rotterdam Study, in silico[22]
USA	275	271	Illumina 550K	NIH	NIH, publicly available[9]
Ireland	221	211	Illumina 550K	NIH	Beaumont Hospital, Dublin, publicly available[7]
Sweden	476	486	Illumina 370K	deCODE	Umeå University Hospital; genotyped
Belgium	305	341	Illumina 370K	deCODE	University Hospital Gasthuisberg; genotyped
Replication phase:					
The Netherlands	108	-	KASPar	Kbiosciences	UMC Utrecht; genotyped
The Netherlands	-	1,935	Illumina 370K	deCODE	RUNMC; in silico[23]
USA	666	672	Illumina 317K	MGH	MGH & Atlanta, previously published[10]
UK	212	237	Illumina 317K	MGH	Kings College London, , previously published[10]
France	230	709	Illumina 317K	Evry	Evry, previously published[10]
Ireland	103	128	Illumina 610K	deCODE	Beaumont Hospital, Dublin, genotyped
Poland	183	544	KASPar	Kbiosciences	Jagiellonian University, Krakow; genotyped
Germany	1,030	-	KASPar	Kbiosciences	Ulm & Berlin; genotyped
Germany	-	1,364	Illumina 550K	Multiple sites*	PopGen, HNR & KORA studies, in silico[24-27]
Germany	-	351	Illumina 317K	Duke University	Munich, in silico[28]

In total data from 2,323 patients and 9,013 controls was included in the genome-wide phase of this study. Data from 2,532 patients and 5,940 controls was available for the 60 SNPs with a p-value 1.0 x x10-4 in the genome-wide phase. UMC Utrecht = University Medical Center Utrecht, The Netherlands. deCODE = deCODE genetics, Reykjavik, Iceland. UCLA = University of California Los Angeles, Center for Neurobehavioral Genetics, Los Angeles, USA. Erasmus MC = Erasmus Medical Center, Rotterdam, The Netherlands. RS-I = ERGO cohort from The Rotterdam Study. NIH = Laboratory of Neurogenetics, National Institute
of Aging, National Institutes of Health, Bethesda, USA. Kbiosciences, Kbiosciences, Hoddesdon, Herts, UK. RUNMC = Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. MGH = Massachusetts General Hospital, Charlestown, USA. Evry = Centre National de Genotypage at Evry. Ulm = Department of Neurology, University of Ulm, Ulm, Germany. Berlin = Department of Neurology, Charité University Hospital, Humboldt-University, Berlin, Germany. *Genotyping experiments for the PopGen, HNR & KORA cohorts were performed at multiple sites and details can be found in the relevant publications. PopGen = Population-based
Recruitment of Patients and Controls for analysis project. HNR = the Heinz Nixdorf Recall Study. KORA = Cooperative Health Research in the Region of Augsburg study. Duke University = Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, United States of America.

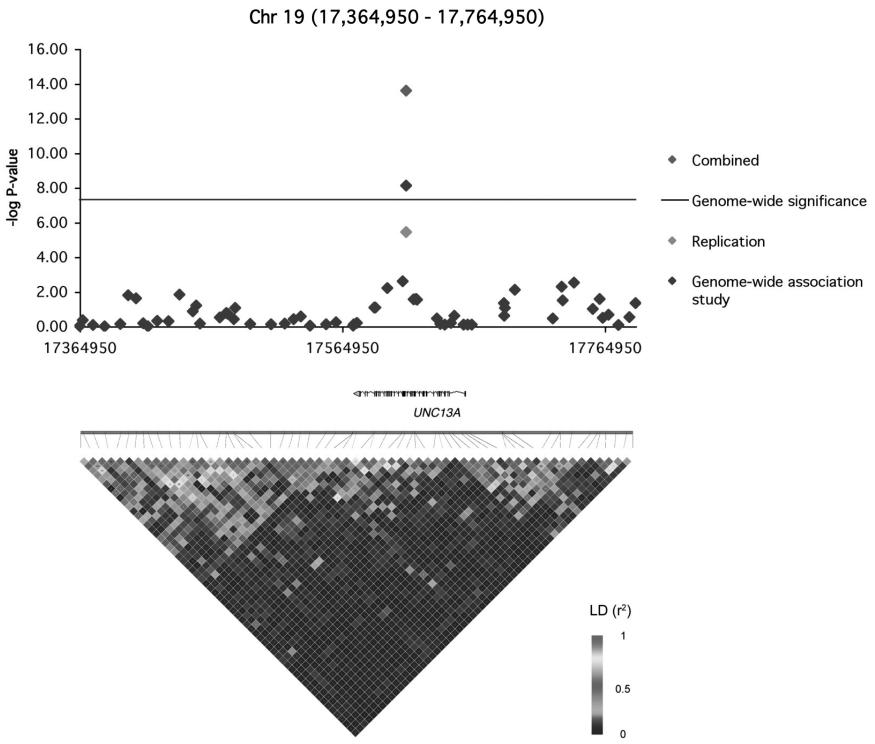


Figure 1. Top part of the figure shows –log10 p-values for all SNPs in a 500 kb region on chromosome 19 surrounding rs12308932. The blue diamonds correspond to the SNPs in the genome-wide association phase. The green diamond shows the –log10 p-values for rs12608932 in the replication phase. The red diamond shows the –log10 p-values for rs12608932 in combined analysis of both stages. Bottom part of the figure shows linkage disequilibrium structure in this region. Colouring in the figure is according to r2. For rs12608932 we observed a genome-wide significant p-value in the genome-wide phase (P = 1.30 x 10-9). This SNP demonstrated robust replication in the second cohort (P = 1.86 x 10-6) and a combined analysis over the two stages yielded a P = 2.53 x 10-14. The association signal maps (using r2 > 0.80) to a 10 kb region within intron 21 of UNC13A, containing only 2 known SNPs (rs12608932 and rs4239633).

we addressed population stratification using PLINK by performing identity-by-state (IBS) analysis. We used PLINK to generate 15,131 unlinked ($r^2 < 0.50$), non-disease associated SNPs across the autosomal chromosomes with a call rate of 100%. Data for these SNPs was then downloaded for the HapMap samples from www.hapmap.org.[13] We then proceeded to calculate the first two dimensions of a multi-dimensional scaling analysis of IBS distances in our samples and the HapMap samples. The first two dimensions for each sample were plotted and non-CEPH individuals were removed (Supplementary Figure 3).

In the replication phase we performed population stratification analysis using 100 randomly selected, unlinked, non-disease associated SNPs across the autosomal

chromosomes with a call rate of 100%. We then reanalyzed the genome-wide data and plotted the first two dimensions of a multi-dimensional scaling analysis of IBS distances based on these 100 SNPs. This demonstrated that we were able to differentiate the included individuals in the genome-wide phase and CEPH samples from the CHB/JPT and YRB samples. These 100 SNPs were subsequently used to analyze population stratification in the replication phase by plotting the first 2 dimensions for each sample and removing non-CEPH samples (Supplementary Figure 4).[13] Association analysis in replication stage was performed identical to that of the GWAS data, using logistic regression adjusting for gender, ancestry (based on the replication data) and nationality (dummy coded).

Three of the sixty SNPs (rs12608932 (at 19p13.3); rs2814707 and rs3849942 (at 9p21.2)) in the replication phase demonstrated significant association after Bonferroni correction at $P < 0.05 / 60 = 8.33 \times 10^{-4}$, of which one (rs12608932) already had yielded independent genome-wide significance in the first stage (Table 2). Considering the fact that p-values in the genome-wide phase and replication phase were corrected for population stratification, we calculate combined p-values by summing weighted z-scores accounting for the respective sizes of the two study phases. Combined p-values for the three SNPs also comfortably satisfy the threshold for genome-wide significance of $P < 5.0 \times 10^{-8}$ (Table 2).[12] The combined p-values for the SNPs, that were previously reported to be associated with ALS, did not approach genome-wide significance (rs10260404 (*DPP6*): $P = 1.5 \times 10^{-3}$; rs6700125 (*FGGY*): $P = 0.11$; rs2306677 (*ITPR2*): $P = 0.08$). Results for all 60 SNPs are provided in Supplementary Table 2.

Table 2. Results for significantly associated SNPs

SNP	Chr	Chr pos	GWAS				Replication				Combined P	Closest gene
			MAF ALS	MAF CON	P	OR	MAF ALS	MAF CON	P	OR		
rs12608932	19	17,613,689	0.40	0.34	1.30×10^{-9}	1.25	0.37	0.34	1.86×10^{-6}	1.20	2.50×10^{-14}	*UNC13A*
rs2814707	9	27,526,397	0.26	0.23	3.33×10^{-6}	1.22	0.26	0.23	2.81×10^{-4}	1.16	7.45×10^{-9}	*MOBKL2B*
rs3849942	9	27,533,281	0.26	0.23	1.58×10^{-6}	1.23	0.26	0.23	7.11×10^{-4}	1.15	1.01×10^{-8}	*C9orf72*

MAF = minor allele frequency. OR = odds ratio. CON = Controls subjects. Results from the genome-wide and replication phase were calculated using logistic regression and combined p-values by summing weighted z-scores.

The strongest association signal was detected for rs12608932 (combined $P = 2.53 \times 10^{-14}$), a common variant in intron 21 of *unc-13 homolog A (C. elegans)* (*UNC13A*) located at chromosome 19p13.11. The region is characterized by a lack of extended linkage disequilibrium (LD) and has not been linked to ALS previously. SNP, rs12608932, is not in significant LD with any other genotyped SNP in this study with maximum of $r^2 = 0.08$ with neighbouring SNPs (Figure 1). Examination of HapMap data shows the presence of only one SNP, rs4239633, in close proximity (10 kb) to rs12608932 with $r^2 > 0.80$ in the CEPH population with the haplotype block well within boundaries of (*UNC13A*). Considering no other genes or microRNAs map to this region, we postulate that *UNC13A* is the gene involved in ALS susceptibility (Supplementary Figure 5).

Members of the UNC13 family, such as UNC13A, are presynaptic proteins found in central and neuromuscular synapses that regulate the release of neurotransmitters, peptides and hormones. Neurotransmitter release at synapses between nerve cells is mediated by calcium-triggered exocytotic fusion of synaptic vesicles with the plasma membrane. Before fusion, vesicles dock at the presynaptic release site, where they are primed to a fusion-competent state. It has been demonstrated that Munc13-1 (murine homologue) is required for synaptic vesicle priming and mice lacking *Munc13-1* have disrupted glutamatergic neurotransmission, due to arrested synaptic vesicle maturation.[14] In addition, *Munc13-1* deficient mice display morphological defects in spinal cord motor neurons, muscle, and neuromuscular synapses.[14] Interestingly, there is accumulating evidence that motor neuron death in ALS is caused by glutamatergic excitotoxicity and the only effective drug for ALS, riluzole, is a glutamate release inhibitor. Thus, it seems plausible that UNC13A may play a role in glutamate mediated excitotoxicity in ALS.[15]

Two SNPs at chromosome 9p21.2 that are in linkage disequilibrium (r^2 = 0.95) also demonstrated significant association in the replication phase and genome-wide significance in the combined analysis (rs2814707: $P = 7.45 \times 10^{-9}$ and rs3849942: $P = 1.01 \times 10^{-8}$) (Table 2). These two SNPs are located in an LD block of approximately 80 kb containing three genes: *MOBKL2B*, *IFNK* and *C9orf72* (Supplementary Figure 6). We fine-mapped this region by genotyping an additional 20 SNPs surrounding rs2814707 and rs3849942, which confirmed the boundaries for the associated LD block (Figure 2.). Haplotype analysis of this locus did not reveal a stronger signal compared to single SNP analysis with $P = 2.76 \times 10^{-6}$ (genome-wide data) and $P = 1.80 \times 10^{-3}$ (replication data) for the haplotype containing associated alleles of rs2814707 and rs3849942 (Supplementary Table 4 and 5). Multivariate logistic regression conditioning the genotypic additive effect of rs12608932 on either rs2814707 or rs3849942, while adjusting for the covariates as in the main analysis, showed that effects of both loci were independent (Supplementary Table 6). In addition, analysis of interaction (defined as a departure from the multiplicative model) between the significantly associated SNPs showed no significant evidence for epistasis (P = 0.11; Supplementary Table 7).

There is increasing evidence that ALS is part of a spectrum with other neurodegenerative disorders such as fronto-temporal dementia (FTD) and Parkinson's disease. Approximately 5% of ALS patients also develop FTD and up to 50% of patients demonstrate cognitive impairment characteristic of frontal dysfunction.[16, 17] Linkage to an 11 Mb region at chromosome 9p13.2-21.3 containing 103 genes has been reported in several families with autosomal dominant ALS with FTD.[18-20] To date, the causal mutation at this locus has not been identified. The significant SNPs in this study (rs2814707 and rs3849942) are located within this locus and tag a haplotype-block of approximately 80 kb. This raises the interesting possibility that variation at this locus may underlie familial as well as sporadic ALS, and potentially FTD, suggesting a more general role of this locus in neurodegeneration.

In summary, our two-stage GWA study has identified two novel loci on 9p21.2 and 19p13.11 that are associated with susceptibility to sporadic ALS. The identification of *UNC13A* as a susceptibility gene suggests that defects in presynaptic machinery and function may contribute to the pathogenic mechanism involved in ALS. The identification of common variants that are robustly associated with susceptibility to sporadic ALS is an exciting step forward in understanding the complex genetics of this devastating disease.

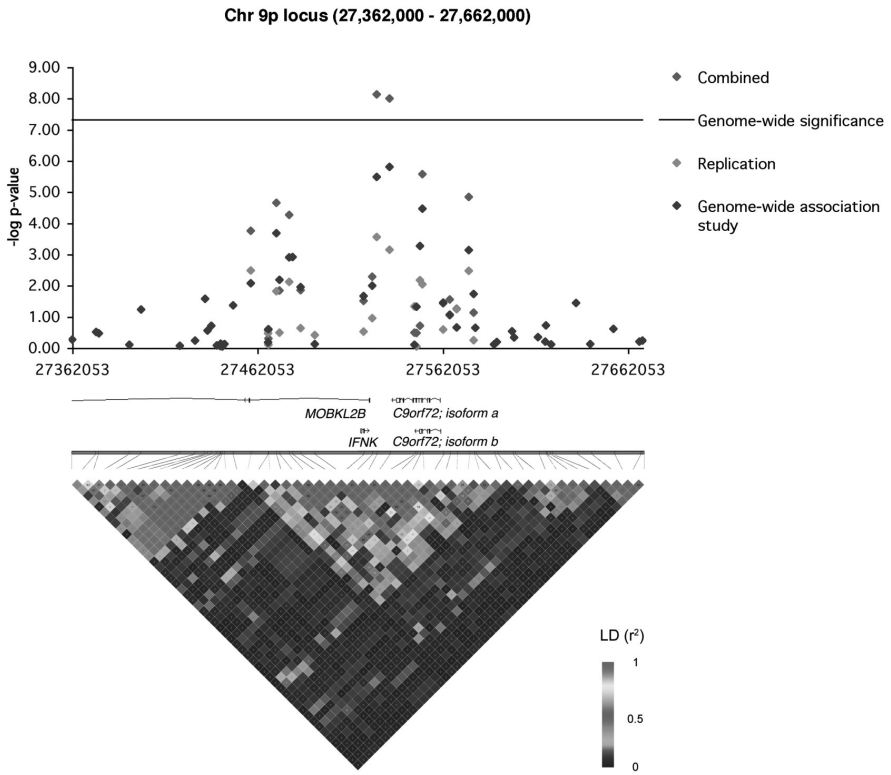


Figure 2. Top part of the figure shows –log10 p-values for all SNPs in a 300 kb region on chromosome 9 surrounding rs2814707 and rs3849942.The blue diamonds correspond to the SNPs in the genome-wide association phase. The green diamonds show the –log10 p-values for the 20 additionally genotyped SNPs in this region in the replication phase. The red diamonds show the –log10 p-values for these SNPs in combined analysis of both stages. Bottom part of the figure shows linkage disequilibrium structure in this region. Colouring in the figure is according to r2. Three SNPs in this region fulfilled the criteria for follow-up in the replication phase; rs2814707, rs3849942 and rs774359. In combined analysis the p-values for rs2814707 and rs3849942 exceeded the threshold of $P = 5.0 \times 10^{-8}$ for genome-wide significance ($rs2814707: P = 7.45 \times 10^{-9}$ and $rs3849942: P = 1.01 \times 10^{-8}$). The association signal maps (using r2 > 0.80) to an LD block of approximately 80 kb containing 3 genes; MOBKL2B, IFNK and C9orf72.

METHODS

Subjects and data. In order to maximize the power of the study, we included data from previously published and publicly available genome-wide association studies in ALS,[4-10] as well as data from newly genotyped subjects. Additional data from healthy control subjects was obtained from different previously published genome-wide association studies.[21-28] All participants gave written informed consent and approval was obtained from the local, relevant ethical committees for medical research. All patients were diagnosed according to the 1994 El Escorial criteria and fulfilled the criteria for probable or definite sporadic ALS.[29] Detailed information on the study population is provided in the Supplementary Information.

DNA. *Dutch samples*: Blood samples were collected in 10 ml EDTA tubes. DNA was isolated from whole blood using the autopure DNA isolation protocol from Qiagen (Qiagen, Valencia, USA). *Belgian samples*: Genomic DNA was extracted from peripheral lymphocytes using standard procedures, on a Chemagen Magnetic Separation Module 1 platform (Chemagen AG, Baesweiler, Germany). *Swedish samples*: Genomic DNA was extracted from white blood cells from peripheral blood using the QIAamp Blood Kit (Qiagen). *German and Polish samples*: DNA was extracted from Epstein-Barr virus immortalized lymphocyte cell lines using the QIAamp Blood Kit (Qiagen).

Details for DNA extraction methods for samples from other genome-wide association studies are provided in the relevant publications. Because only a limited quantity of DNA was available from the Polish controls, these samples were amplified. DNA amplification was performed using REPLI-g Mini amplification kits for whole genome amplification from Qiagen according to the manufacturer's protocol with 5 ng genomic as input.

Genotyping procedures. In the genome-wide phase all experiments were carried out using high-density, high-throughput genotyping Beadchips manufactured by Illumina (Illumina, San Diego, USA). Data from multiple studies and laboratories was included in this study and data was generated using different versions of Illumina Beadchips. An overview of genotyping facilities and chip types is provided in Table 1 of the main article.

In the replication phase samples were either genotyped on Illumina Beadchips or using KASPar genotyping technology. All KASPar assays were performed by KBiosciences, Hoddesdon, Herts, UK. All experiments were performed according to the manufacturer's protocol.

Although different versions of the Beadchips were used, the workflow for all chips is very similar. In short, 750 ng of DNA per sample is whole-genome amplified, fragmented, precipitated and resuspended in the appropriate hybridization buffer. Subsequently denatured samples are hybridized on Illumina BeadChips at 48°C for a minimum of 16 hours. After hybridization, the BeadChips are processed for single base extension reactions and stained. Chips are then imaged using the Illumina Bead Array Readers. For each sample, normalized bead intensity data was loaded into Illumina Beadstudio and converted into genotypes. Genotypes were called using the auto-calling algorithm in Illumina Beadstudio.

KASPar is a PCR based genotyping method combined with allele-specific amplification followed by fluorescence detection.

Quality Control. Quality control was performed on each data set separately using PLINK v1.05 (http://pngu.mgh.harvard.edu/~purcell/plink/).[11] In each cohort we excluded all samples and SNPs with a call rate < 95%. SNPs for which deviation from the Hardy-Weinberg Equilibrium was observed with $P < 1.0 \times 10^{-4}$ in controls were also excluded from further analysis. All samples for which the genotypic gender did not match the gender in the phenotype files were removed from the data sets. Files were then merged one by one into one large file for the GWAS. PLINK was used to flip strands when tri-allelic SNPs were detected and after merging two files a flip-scan was performed. An overview is provided in Supplementary Table 1. For all SNPs with $P < 1.0 \times 10^{-4}$, the cluster plots were visually inspected in Beadstudio and the Illumina probe sequence were checked using the Blat function in the UCSC genome browser (http://genome.ucsc.edu/). Similar quality control was performed on the replication data, the only difference being that SNPs with a call rate < 90% were excluded from analysis (Supplementary Table 3).

To assess whether the KASPar and Illumina platforms yielded identical genotypes for the

SNPs in this study we genotyped 152 samples from the GWAS using KASPar technology. Concordance between the two platforms was > 99.9 %. Because we used amplified DNA for the Polish control cohort, we also compared the performance of amplified DNA to genomic DNA on the KASPar platform. 96 samples for which adequate genomic DNA was available were amplified. Both the amplified and genomic samples were genotyped using KASPar. Concordance between the amplified and genomic samples was > 99.9 %. Five samples were genotyped in duplo using the Illumina Beadchips and demonstrated a concordance > 99.9 %.

SNPs. The vast majority of samples (91.4%) in this study were genotyped on Illumina 300K or 370K Beadchips. Only 978 out of 11,336 samples were genotyped on the Illumina 550K SNP arrays. Considering all included samples were from European descent and coverage of the Illumina 300K Beadchips is considered adequate in Caucasian populations we chose not to use imputation. We included all SNPs that passed QC and that were common to all datasets (292,768 SNPs).

Statistical analysis of SNP data. Association testing was performed with the statistical analysis program PLINK v1.05[11] using logistic regression adjusting for gender, ancestry and nationality (dummy coded). For ancestry, we used the first 2 dimensions of the multidimensional scaling on the matrix of identity-by-state sharing of all pairs of individuals as covariates. We then plotted the observed –log10 p-values against the expectation under the null hypothesis. The qq-plot and genomic control λ were calculated using the statistical analysis program R.

Considering the fact that p-values in the genome-wide phase and replication phase were corrected for population stratification, we determined combined p-values calculating total weighted z-scores, accounting for the respective sizes of the two study phases.[30] Results were considered significant when combined p-values were smaller than the proposed threshold for genome-wide significance of $P = 5.0 \times 10^{-8}$.[12] Results for all SNPs that were analyzed in the replication phase are shown in Supplementary Table 2.

We also analyzed the data using a second strategy. We tested whether samples belong to a genetic cluster using a pair-wise population concordance (PPC) test. This is a simple significance test for whether two individuals belong to the same random-mating population. We imposed the constraint on the clustering procedure that all clusters had to contain at least one case and one control. We then performed a Cochran Mantel-Haenszel test on basic allele counts for all 292,768 included SNPs accounting for each genetic cluster. Results from this analysis demonstrated the same ranking of SNPs compared to the logistic regression and lower p-values for the significantly associated SNPs (Supplementary Table 8).

LD structure surrounding the significantly associated SNPs was analyzed using Haploview (http://www.broad.mit.edu/mpg/haploview/)[31] and SNAP (http://www.broad.mit.edu/mpg/snap/).[32] We plotted p-values for SNPs in 500 kb window surrounding the associated SNPs and compared LD structure in our data to the HapMap samples. This demonstrated that none of the SNPs on the Illumina Beadchips is in linkage disequilibrium with the associated SNP in *UNC13A*. Only one SNP in the HapMap data has an $r^2 > 0.8$ with rs12608932 and the LD for the associated locus does not extend outside the boundaries of *UNC13A* (Supplementary Figure 5).

Analysis of the associated locus on chromosome 9 shows that these SNPs are located in an LD block of approximately 80 kb, which contains 3 genes: *MOBKL2B*, *IFNK* and *C9orf72* (Supplementary Figure 6).

Haplotype analysis of chromosome 9p locus. We performed haplotype analysis on

a 300 kb window surrounding the associated SNPs on chromosome 9 using the data from the genome-wide association studies and using data from the 20 SNPs that we genotyped for the fine-mapping experiments in the replication samples. Haplotype blocks were defined using the confidence interval setting as described by Gabriel et al. as implemented in Haploview.[31]

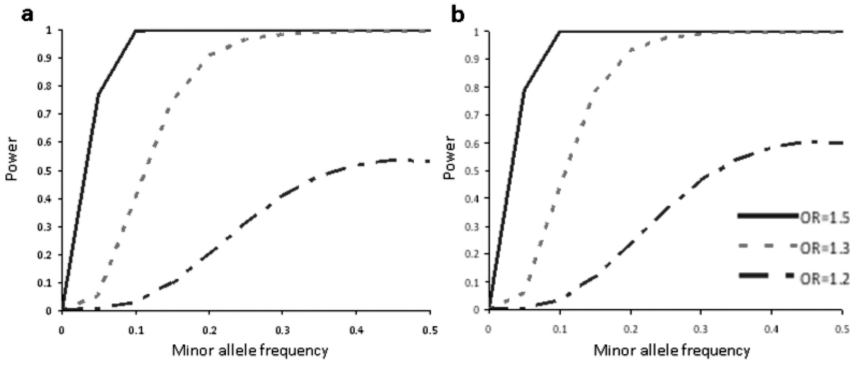
SNP-SNP interaction analysis. We tested if there was evidence for interaction between the significantly associated SNPs in our study by recoding the genotypes to an additive genetic model, with wildtype homozygotes as 0, heterozygotes as 1 and rare homozygotes as 2 and subsequently analyzing the data using the logistic regression, including the covariates as used in the main analysis, and in addition the genotypes of the two SNPs and their interaction. Results are shown in Supplementary Table 7. There is no evidence for significant interaction. We also calculated the interaction between these two SNPs with the genotypes coded as factors in order to detect more complex pattern of interactions, but p-values were all > 0.15.

REFERENCES

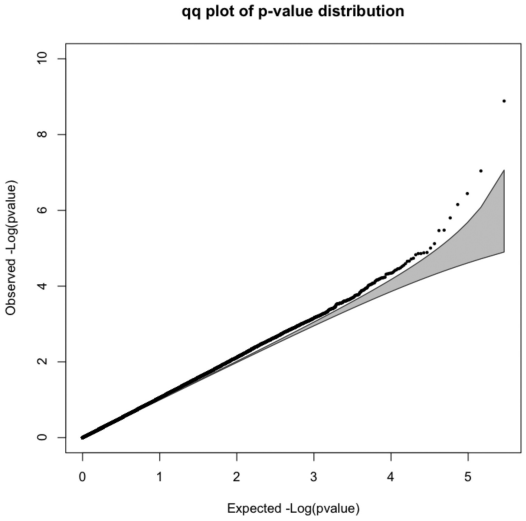
1. Mitchell,J.D. & Borasio,G.D. Amyotrophic lateral sclerosis. *Lancet* 369, 2031-2041 (2007).
2. Miller,R.G., Mitchell,J.D., Lyon,M., & Moore,D.H. Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND). *Cochrane. Database. Syst. Rev.*CD001447 (2007).
3. Graham,A.J., Macdonald,A.M., & Hawkes,C.H. British motor neuron disease twin study. *J. Neurol. Neurosurg. Psychiatry* 62, 562-569 (1997).
4. van Es,M.A. *et al.* Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.* 40, 29-31 (2008).
5. van Es,M.A. *et al.* ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet Neurol.* 6, 869-877 (2007).
6. Dunckley,T. *et al.* Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N. Engl. J. Med.* 357, 775-788 (2007).
7. Cronin,S. *et al.* A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum. Mol. Genet.* 17, 768-774 (2008).
8. Chio,A. *et al.* A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.*(2009).
9. Schymick,J.C. *et al.* Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 6, 322-328 (2007).
10. Landers,J.E. *et al.* Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. U. S. A* 106, 9004-9009 (2009).
11. Purcell,S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559-575 (2007).
12. Pe'er,I., Yelensky,R., Altshuler,D., & Daly,M.J. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.* 32, 381-385 (2008).
13. The International HapMap Project. *Nature* 426, 789-796 (2003).
14. Varoqueaux,F., Sons,M.S., Plomp,J.J., & Brose,N. Aberrant morphology and residual transmitter release at the Munc13-deficient mouse neuromuscular synapse. *Mol. Cell Biol.* 25, 5973-5984 (2005).
15. Augustin,I., Rosenmund,C., Sudhof,T.C., & Brose,N. Munc13-1 is essential for fusion competence of glutamatergic synaptic vesicles. *Nature* 400, 457-461 (1999).
16. Phukan,J., Pender,N.P., & Hardiman,O. Cognitive impairment in amyotrophic lateral sclerosis. *Lancet Neurol.* 6, 994-1003 (2007).
17. Raaphorst,J., De,V.M., Linssen,W.H., De Haan,R.J., & Schmand,B. The cognitive profile of amyotrophic lateral

sclerosis: A meta-analysis. *Amyotroph. Lateral. Scler.*1-13 (2009).
18. Valdmanis,P.N. *et al.* Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p. *Arch. Neurol.* 64, 240-245 (2007).
19. Morita,M. *et al.* A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. *Neurology* 66, 839-844 (2006).
20. Vance,C. *et al.* Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain* 129, 868-876 (2006).
21. Stefansson,H. *et al.* Large recurrent microdeletions associated with schizophrenia. *Nature* 455, 232-236 (2008).
22. Hofman,A. *et al.* The Rotterdam Study: objectives and design update. *Eur. J. Epidemiol.* 22, 819-829 (2007).
23. Kiemeney,L.A. *et al.* Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.* 40, 1307-1312 (2008).
24. Krawczak,M. *et al.* PopGen: population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet.* 9, 55-61 (2006).
25. Wichmann,H.E., Gieger,C., & Illig,T. KORA-gen--resource for population genetics, controls and a broad spectrum of disease phenotypes. *Gesundheitswesen* 67 Suppl 1, S26-S30 (2005).
26. Schmermund,A. *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am. Heart J.* 144, 212-218 (2002).
27. Luca,D. *et al.* On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *Am. J. Hum. Genet.* 82, 453-463 (2008).
28. Need,A.C. *et al.* A genome-wide investigation of SNPs and CNVs in schizophrenia. *PLoS. Genet.* 5, e1000373 (2009).
29. Brooks,B.R. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J. Neurol. Sci.* 124 Suppl, 96-107 (1994).
30. de Bakker,P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* 17, R122-R128 (2008).
31. Barrett,J.C., Fry,B., Maller,J., & Daly,M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 21, 263-265 (2005).
32. Johnson,A.D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 24, 2938-2939 (2008).

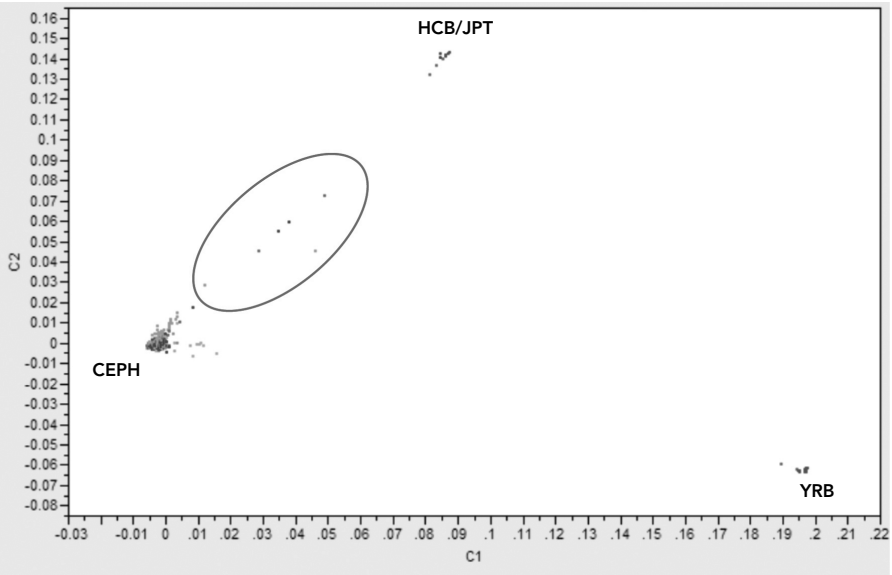
SUPPLEMENTARY MATERIAL



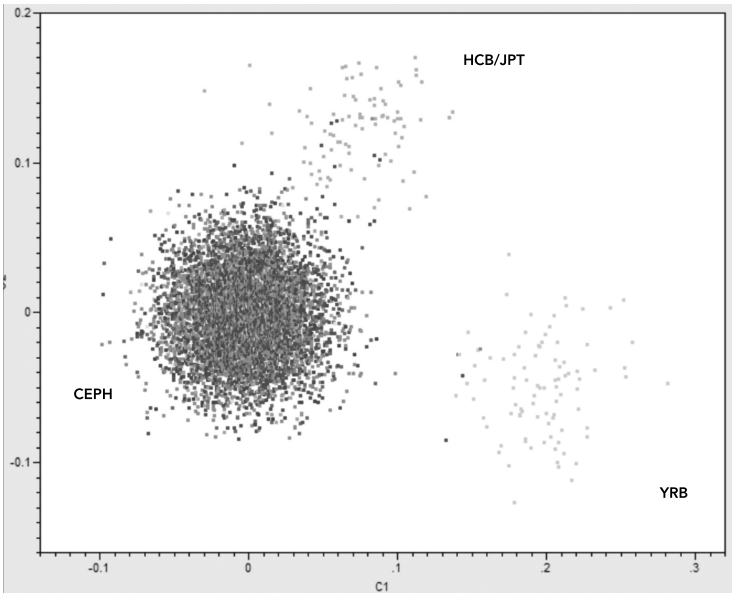
Supplementary Figure 1. *Power to detect a genetic association at P = 5.0 x 10-8 for the genome-wide association study (a) and the replication phase (b) is shown. Study was designed to have equal power in both stages.*



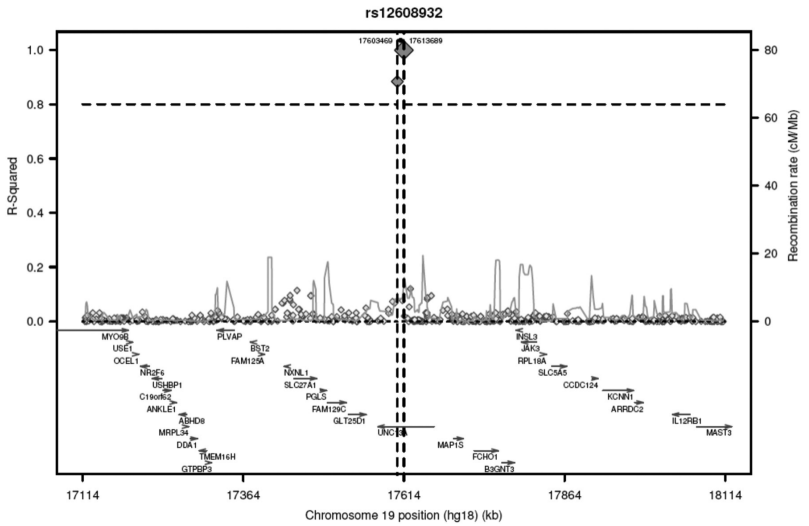
Supplementary Figure 2. *A quantile-quantile plot of observed –log10 P versus the expectation under the null for the genome-wide data. The figure shows departure from the null with GC = 1.05.*



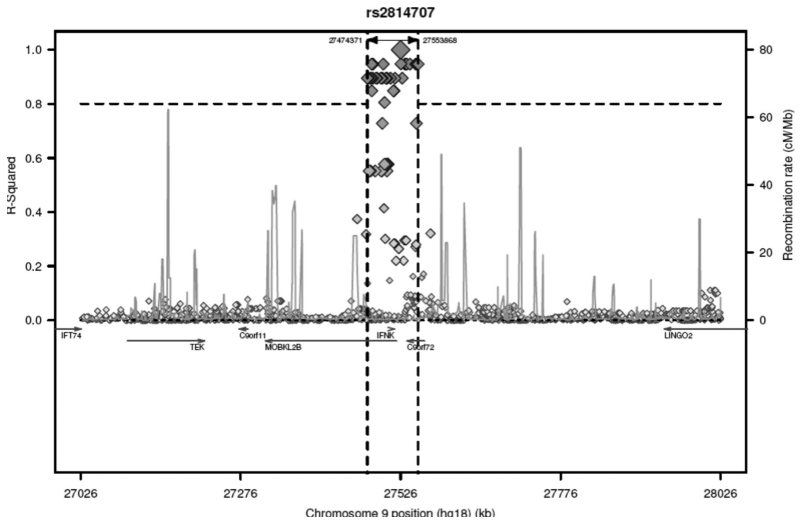
Supplementary Figure 3. A plot of the first two dimensions (C1 and C2) of a multi-dimensional scaling analysis of IBS distances in study samples and the HapMap samples is shown. The samples in the red circle were removed as well the study samples (marked in red) in the HCB/JPT (n = 3) and YRB (n = 1) clusters.



Supplementary Figure 4. A plot of the first two dimensions (C1 and C2) of a multi-dimensional scaling analysis of IBS distances in study samples and the HapMap samples calculated using 100 SNPs in the replication data.



Supplementary Figure 5. *SNAP plot showing an overview of all SNPs with an r2 > 0.80 with rs12608932. The diamonds in the figure represent SNPs. The top SNP in the plot is rs12608932. On the x-axis corresponding chromosomal positions are plotted for each SNP. On the y-axis (left side) r2-values for linkage disequilibrium for each SNP with the associated SNP (rs12608932) is shown. The y-axis on the right side shows estimated recombination rates in centimorgans per megabase from the HapMap (v21) Phase II data. The vertical dotted lines mark the extent of linkage disequilibrium for the associated region. The plot clearly shows that only one SNP, rs4239633 within a distance of 10 kb, has an r2-value > 0.80 with the associated SNP. Linkage disequilibrium does not extend beyond the boundaries of UNC13A.*



Supplementary Figure 6. SNAP plot showing an overview of all SNPs with an r2 > 0.80 with rs2814707. The diamonds in the figure represent SNPs. The top SNP in the plot is rs2814707. On the x-axis corresponding chromosomal positions are plotted for each SNP. On the y-axis (left side) r2-values for linkage disequilibrium for each SNP with the associated SNP (rs2814707) is shown. The y-axis on the right side shows estimated recombination rates in centimorgans per megabase from the HapMap (v21) Phase II data. The vertical dotted lines mark the extent of linkage disequilibrium for the associated region. The figure shows that rs2814707 and rs3849942 map to an LD block of approximately 80kb containing 3 genes: MOBKL2B, IFNK and C9orf72.

Supplementary Table 1. Summary statistics for quality control on genome-wide data

	Pre QC		Filters					Post QC			
Country	SNPs	Samples	Samples call rate <95%	Failed gender check	SNPs MAF <5%	HWE P < 1.0 x 10^{-4}	SNPs call rate <95%	Cases	Controls	SNPs	Call rate (%)
Belgium	311,995	681	27	8	9,084	159	1,646	305	341	301,382	99.45
USA	311,995	547	1	0	11,616	162	5,106	275	271	298,195	98.60
Sweden	311,995	973	0	11	10,505	217	941	476	486	300,595	99.61
Netherlands	311,995	8,818	14	54	8,298	3,403	3,075	1,046	7,704	298,279	99.57
Ireland	311,995	432	0	0	9,695	182	1,606	221	211	300,077	99.76
Total	311,995	11,451	42	73	13,541	4,123	12,374	2,323	9,013	292,768	99.79

Table shows the number of SNPs common to all datasets prior to QC, the number of samples per population, the number of samples with a call rate < 95%, the number of samples for which there was a discrepancy between the sex in the phenotype file and the genotypic gender, the number of SNPs with a minor allele frequency (MAF) < 5%, the number of SNPs for which a deviation from the Hardy Weinberg Equilibrium was observed with $P < 1.0 \times 10^{-4}$, the number of SNPs with a call rate < 95% and the number of cases, controls, SNPs and average call rate remaining after quality control.

Supplementary Table 2. Results for all SNPs with P < 1.0 x 10^{-4} in GWAS

SNPs			GWAS				Replication				Combined
SNP	Chr	Chr pos	MAF Case	MAF Con	*P*	OR	MAF Case	MAF Con	*P*	OR	*P*
rs12608932	19	17,613,689	0.40	0.34	1.30 x 10^{-9}	1.25	0.37	0.34	1.86 x 10^{-6}	1.20	2.53 x 10^{-14}
rs2814707	9	27,526,397	0.26	0.23	3.33 x 10^{-6}	1.22	0.26	0.23	2.81 x 10^{-4}	1.16	7.45 x 10^{-9}
rs3849942	9	27,533,281	0.26	0.23	1.58 x 10^{-6}	1.23	0.26	0.23	7.11 x 10^{-4}	1.15	1.01 x 10^{-8}
rs5937496	X	75,263,826	0.17	0.13	9.12 x 10^{-8}	1.38	0.16	0.16	0.14	1.13	5.96 x 10^{-7}
rs9971637	12	17,265,006	0.06	0.05	3.60 x 10^{-7}	1.48	0.05	0.05	0.09	1.16	1.58 x 10^{-6}
rs774359	9	27,551,049	0.28	0.25	3.50 x 10^{-5}	1.19	0.27	0.25	0.01	1.11	2.69 x 10^{-6}
rs2405657	11	97,187,774	0.32	0.34	1.38 x 10^{-5}	0.84	0.33	0.35	0.02	0.91	3.19 x 10^{-6}
rs5916687	X	4,606,138	0.29	0.27	4.14 x 10^{-5}	1.22	0.3	0.29	0.01	1.14	3.36 x 10^{-6}
rs3734399	6	11,246,024	0.17	0.15	6.10 x 10^{-6}	1.22	0.17	0.16	0.98	1.00	1.08 x 10^{-5}
rs4792192	17	11,744,186	0.09	0.10	7.27 x 10^{-5}	0.78	0.10	0.11	0.06	0.89	5.05 x 10^{-5}
rs1396919	2	115,544,258	0.18	0.15	6.85 x 10^{-5}	1.21	0.17	0.15	0.07	1.10	5.96 x 10^{-5}
rs395119	9	131,035,892	0.29	0.27	1.84 x10^{-5}	1.19	0.28	0.27	0.14	1.06	6.47 x 10^{-5}
rs4944178	11	77,270,869	0.08	0.10	3.41 x 10^{-6}	0.73	0.1	0.1	0.44	0.94	9.04 x 10^{-5}
rs8062856	16	67,640,318	0.49	0.45	6.49 x 10^{-5}	1.16	0.47	0.47	0.90	1.00	9.55 x 10^{-5}
rs613479	15	32,146,482	0.34	0.35	6.21 x 10^{-5}	0.86	0.35	0.36	0.11	0.94	1.07 x 10^{-4}
rs917379	5	129,485,823	0.12	0.11	8.36 x 10^{-5}	1.25	0.12	0.12	0.89	1.01	1.32 x 10^{-4}
rs823538	2	2,472,777	0.47	0.44	9.90 x 10^{-6}	1.18	0.47	0.46	0.74	1.01	1.34 x 10^{-4}
rs5768827	22	45,266,582	0.37	0.34	6.90 x 10^{-5}	1.16	0.33	0.33	0.87	1.01	1.48 x 10^{-4}
rs1989	15	81,572,635	0.08	0.09	6.01 x 10^{-5}	0.76	0.1	0.09	0.83	0.99	2.07 x 10^{-4}
rs6431612	2	239,175,763	0.29	0.31	1.93 x 10^{-5}	0.84	0.31	0.31	0.28	0.96	2.42 x 10^{-4}
rs9327807	5	137,378,377	0.15	0.18	2.22 x 10^{-5}	0.81	0.18	0.18	0.69	0.98	3.17 x 10^{-4}
rs9393338	6	22,935,562	0.06	0.07	8.18 x 10^{-5}	0.73	0.06	0.06	0.78	0.98	4.00 x 10^{-4}
rs10079678	5	129,534,547	0.09	0.07	8.28 x 10^{-5}	1.3	0.08	0.08	0.76	1.02	4.70 x 10^{-4}
rs2045333	1	2,791,319	0.13	0.10	4.77 x 10^{-5}	1.26	0.10	0.10	0.30	1.07	4.78 x 10^{-4}
rs1776148	1	240,109,168	0.39	0.37	1.38 x 10^{-5}	1.18	0.37	0.37	0.45	1.03	5.83 x 10^{-4}
rs7956931	12	18,273,654	0.16	0.18	4.75 x 10^{-5}	0.82	0.17	0.17	0.65	0.98	6.81 x 10^{-4}
rs450652	9	89,013,833	0.07	0.09	7.56 x 10^{-6}	0.73	0.08	0.08	0.56	1.04	7.62 x 10^{-4}
rs11007742	10	30,116,332	0.12	0.10	1.48 x 10^{-5}	1.28	0.11	0.11	0.49	1.04	7.63 x 10^{-4}
rs201074	10	11,037,319	0.39	0.42	9.33 x 10^{-5}	0.86	0.4	0.41	0.31	0.96	8.24 x 10^{-4}
rs758384	17	6,100,103	0.25	0.23	4.42 x 10^{-5}	1.19	0.23	0.23	0.4	1.04	8.82 x 10^{-4}
rs7565870	2	51,884,261	0.23	0.25	6.50 x 10^{-5}	0.84	0.24	0.24	0.38	0.96	9.95 x 10^{-4}
rs17643851	8	18,482,009	0.05	0.07	7.02 x 10^{-7}	0.66	0.05	0.05	0.19	1.12	0.0011
rs10489650	1	79,503,341	0.33	0.30	3.45 x 10^{-5}	1.17	0.29	0.29	0.49	0.97	0.0013
rs2014307	10	124,207,622	0.42	0.40	5.88 x 10^{-5}	1.16	0.41	0.40	0.56	1.02	0.0013
rs10260404	7	153,841,731	0.41	0.38	9.23 x 10^{-5}	1.16	0.37	0.39	0.42	0.97	0.0016
rs350729	2	52,837,277	0.34	0.31	7.47 x 10^{-5}	1.17	0.33	0.33	0.51	1.03	0.0019
rs1129212	16	86,289,379	0.12	0.10	9.86 x 10^{-5}	1.25	0.10	0.11	0.60	0.96	0.002
rs4266132	3	138,693,225	0.20	0.18	5.82 x 10^{-5}	1.21	0.19	0.20	0.57	0.97	0.0026
rs11264236	1	151,925,104	0.32	0.30	1.30 x 10^{-5}	1.19	0.30	0.31	0.25	0.96	0.0031
rs2723708	4	169,829,244	0.44	0.46	1.32 x 10^{-5}	0.85	0.45	0.45	0.75	1.01	0.0031
rs12162045	16	23,239,244	0.28	0.25	9.17 x 10^{-5}	1.17	0.25	0.26	0.41	0.97	0.0037
rs310819	12	76,016,205	0.35	0.33	3.27 x 10^{-5}	1.17	0.34	0.34	0.70	0.99	0.0037
rs137514	22	31,616,796	0.06	0.09	4.63 x 10^{-5}	0.74	0.08	0.08	0.84	1.02	0.0043

rs7161223	14	95,196,835	0.35	0.38	4.58×10^{-5}	0.86	0.39	0.37	0.29	1.04	0.0048
rs3817287	10	124,086,377	0.33	0.30	8.87×10^{-5}	1.17	0.31	0.32	0.30	0.96	0.0064
rs1527047	12	76,297,680	0.35	0.37	8.85×10^{-5}	0.86	0.38	0.37	0.74	1.01	0.0079
rs11159460	14	79,942,688	0.28	0.29	3.84×10^{-5}	0.85	0.32	0.31	0.82	1.01	0.0084
rs6729829	2	81,341,196	0.27	0.25	6.02×10^{-5}	1.18	0.25	0.26	0.20	0.95	0.0095
rs938111	3	118,330,571	0.13	0.11	6.83×10^{-5}	1.25	0.13	0.13	0.81	0.99	0.01
rs3757155	6	136,500,286	0.37	0.34	3.01×10^{-5}	1.17	0.36	0.37	0.13	0.94	0.01
rs2046091	6	136,500,942	0.40	0.37	9.49×10^{-5}	1.16	0.39	0.40	0.19	0.95	0.01
rs3799386	6	136,491,840	0.37	0.34	4.53×10^{-5}	1.17	0.36	0.37	0.12	0.94	0.01
rs6759238	2	218,244,694	0.40	0.43	2.55×10^{-5}	0.86	0.45	0.44	0.09	1.06	0.01
rs10460926	3	82,570,758	0.15	0.17	2.18×10^{-5}	0.8	0.19	0.18	0.52	1.04	0.01
rs11595807	10	21,808,932	0.10	0.09	2.80×10^{-5}	1.29	0.09	0.10	0.08	0.90	0.02
rs17429690	1	5,736,722	0.13	0.12	8.41×10^{-5}	1.24	0.10	0.11	0.13	0.91	0.02
rs1706871	2	2,451,162	0.51	0.48	4.89×10^{-5}	1.16	0.50	0.50	0.94	1.00	0.03
rs7193938	16	52,590,636	0.20	0.18	3.72×10^{-5}	1.21	0.17	0.19	0.02	0.89	0.05
rs2306677	12	26,527,653	0.10	0.09	0.06	1.13	0.09	0.09	0.61	1.03	0.08
rs6700125	1	59,475,385	0.34	0.33	0.43	1.03	0.36	0.34	0.05	1.08	0.11

Table shows results for all SNPs with a p-value < 1.0×10^{-4} in the genome-wide phase. MAF = minor allele frequency. P-values were calculated with PLINK using logistic regression with gender, the first 2 dimensions from the multi-dimensional scaling analysis of IBS distances and nationality (dummy coded) as covariates. Combined p-values were calculated by summing weighted z-scores over the two stages using R.

Supplementary Table 3. Summary statistics for quality control on replication data

	Pre QC		Filters					Post QC			
Country	**SNPs**	**Samples**	**Samples call rate <90%**	**Failed gender check**	**SNP MAF <5%**	**HWE P <1x10^{-4}**	**SNPs call rate <90%**	**Cases**	**Controls**	**SNPs**	**Call rate (%)**
USA	160	1,338	0	0	0	0	0	666	672	160	99.69
France	160	939	0	0	0	0	0	230	709	160	99.82
UK	160	449	0	0	0	0	0	212	237	160	99.09
Netherlands Controls	160	1,980	45	0	0	0	0	-	1,935	160	99.75
Netherlands ALS	160	121	13	0	0	0	0	108	-	160	99.11
Germany ALS	160	1,051	21	0	0	0	0	1,03	-	160	99.03
Germany Controls1	160	1,364	0	0	0	0	0	-	1,364	160	99.35
Germany Controls2	160	351	0	0	0	0	0	-	351	160	98.99
Ireland	160	231	0	0	0	0	0	103	128	160	99.98
Poland	160	727	65	0	0	0	0	183	544	160	98.75
Total	160	8,551	144	0	0	0	0	2,532	5,940	160	99.44

Table shows the number of SNPs common to all datasets prior to QC, the number of samples per population, the number of samples with a call rate < 90%, the number of samples for which there was a discrepancy between the sex in the phenotype file and the genotypic gender, the number of SNPs with a minor allele frequency (MAF) < 5%, the number of SNPs for which a deviation from the Hardy Weinberg Equilibrium was observed with $P < 1.0 \times 10^{-4}$, the number of SNPs with a call rate < 90% and the number of cases, controls, SNPs and average call rate remaining after quality control.

Supplementary Table 4. Results from haplotype analysis in genome-wide data

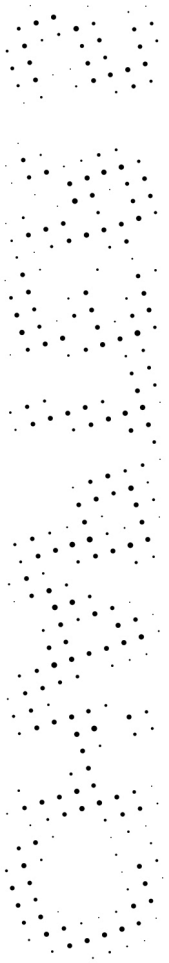
Haplotypes	**Frequency cases**	**Frequency controls**	***P***
CGC**AA**GA	0.26	0.22	2.76×10^{-6}
CAAGGGG	0.22	0.23	0.14
CGCGGAA	0.21	0.21	0.49
CAAGGGA	0.15	0.15	0.85
AAAGGGA	0.09	0.09	0.62
CGCGGGA	0.04	0.04	0.03
CGAGGGA	0.02	0.02	0.11
CGCGGGG	0.01	0.01	0.99

The associated alleles of rs2814707 and rs3849942 are marked bold in the first haplotype. Haplotype definition was performed using the confidence interval method as described by Gabriel et al. in Haploview.

Supplementary Table 5. Results from haplotype analysis in replication data

Haplotypes	Frequency cases	Frequency controls	*P*
CGC**AA**GA	0.25	0.23	1.80×10^{-3}
CAAGGGG	0.21	0.21	0.64
CGCGGAA	0.19	0.21	0.01
CAAGGGA	0.16	0.16	0.20
AAAGGGA	0.10	0.10	0.55
CGCGGGA	0.05	0.05	0.57
CGAGGGA	0.02	0.02	0.06
CGCGGGG	0.02	0.01	0.27

The associated alleles of rs2814707 and rs3849942 are marked bold in the first haplotype. Haplotype definition was performed using the confidence interval method as described by Gabriel et al. in Haploview.



COPY-NUMBER VARIATION IN SPORADIC AMYOTROPHIC LATERAL SCLEROSIS: A GENOME-WIDE SCREEN

The Lancet Neurology (2008)

Hylke M Blauw[1,6], Jan H Veldink[1,6], Michael A van Es[1], Paul W van Vught[1], Christiaan GJ Saris[1], Bert van der Zwaag[2], Lude Franke[3], J Peter H Burbach[2], John JH Wokke[1], Roel A Ophoff[4,5] & Leonard H van den Berg[1]

[1]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [2]Department of Pharmacology and Anatomy, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [3]Complex Genetics Section, Department of Biomedical Genetics, University Medical Center Utrecht, The Netherlands; [4]Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [5]Neuropsychiatric Institute and Department of Human Genetics, University of California, Los Angeles, USA; [6]These authors contributed equally to this work

SUMMARY

Background: Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease characterised by the selective death of motor neurons in the brain and spinal cord. Genetic risk factors are considered to be implicated in disease susceptibility. In addition to single nucleotide polymorphisms (SNPs), copy-number variants (CNVs) are a newly recognized source of genetic variation with important impact on gene expression and disease phenotypes. Our aim was to identify CNVs that predispose to sporadic ALS.

Methods: We did a genome-wide screen for CNVs by analysing Illumina 317K SNP arrays in a large cohort of ALS patients and matched controls (n=406 and 404 respectively). We examined CNVs for association with ALS and further analysed genes that were exclusively deleted in ALS patients.

Findings: We detected 2,328 CNVs in 810 individuals. No CNV locus was significantly associated with ALS. 406 genes were duplicated or deleted exclusively in ALS patients and have not been reported in CNVs previously. Of the 390 genes heterozygously deleted in ALS patients, 155 (40%) were exclusively observed in ALS patients. In contrast, of the 323 genes heterozygously deleted in controls, only 51 (16%) were exclusive to the controls, which was statistically significant ($p=2{\cdot}15x10^{-12}$). The ALS-specific deleted genes represented biological functions including oxidative phosphorylation, regulation of actin cytoskeleton and cytokine-cytokine receptor interaction.

Interpretation: We show that common CNVs in the regions of the genome represented on the SNP array are unlikely to be associated with ALS. However, the high number of ALS-specific deleted genes strongly suggests that multiple rare deletions may play an important role in ALS pathogenesis.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS), a form of motor neuron disease, is a fatal neurodegenerative disease characterised by the selective loss of motor neurons in the cortex, brainstem and spinal cord. Patients suffer from progressive wasting and weakness of limb, bulbar and respiratory muscles, and die on average within 3 years of symptom onset, usually because of respiratory failure[1]. Although ALS can occur anytime during adulthood, the median age of onset is in the mid-fifties. The only therapeutic strategy to slow progression of ALS is currently riluzole, which delays disease development by 3 to 6 months[2]. Global incidence of ALS is about 1-2 per 100,000 and lifetime risk of developing ALS is estimated to be 1/600-1/2000, which makes it the most common motor neuron disease[1,3].

About 10% of patients have a family history of ALS (familial ALS). The remaining 90% of ALS cases are sporadic[3], and are thought to be multifactorial, with both environmental and genetic components contributing to disease susceptibility[4]. On the basis of concordance rates in twin studies, estimates of the heritability of ALS range from 0·38 to 0·85[5]. The search for pathogenic genetic variants has primarily focused on variations at the single nucleotide polymorphism (SNP) level. Using a candidate gene or genome-wide approach, SNPs have been identified that may increase susceptibility for ALS including SNPs in *VEGF*, *ANG, FLJ10986*, *ITPR2*, and *DPP6*[6-10].

In addition to variation at the SNP level, genomic copy-number variants (CNVs) constitute an important source of genetic and phenotypic variation[11]. Copy-number variation is caused by chromosomal rearrangements that result in the loss (i.e. deletions) or gain (i.e. duplications) of stretches of DNA sequence. Recent publications have shown that CNVs are ubiquitous in the human genome and encompass more nucleotide content than the total number of single nucleotide polymorphisms (SNPs), until recently considered to be the predominant form of genetic variation[11]. CNVs are typically 1 kilobase to several megabases in size and cannot be detected by conventional karyotyping techniques. Currently, over 3,600 known CNV loci covering >18% of the human genome have been identified in healthy controls (Database of Genomic Variants, http://projects.tcag.ca/variation). CNVs have an important impact on gene expression[12], resulting from a dose-effect for genes that are encompassed by a deletion or duplication, or by altered expression through disruption of regulatory elements[13]. Several CNVs have been related to diseases such as Charcot-Marie-Tooth disease and spinal muscular atrophy (SMA), but also complex disease traits such as Crohn's disease, SLE glomerulonephritis, psoriasis and HIV susceptibility[14-18].

Indeed, the mechanism of copy-number variation that changes gene-dose rather than gene-function is in principle compatible with late-onset, complex traits such as ALS[19]. Using a candidate gene approach, we showed previously that variation in copy-numbers of the Survival of Motor Neuron gene (*SMN1* and *SMN2*) is associated with the severity of and susceptibility to ALS in two independent Dutch ALS populations[20]. A low *SMN* gene copy number was associated with ALS affection status, and genotypes producing less SMN protein were more prevalent in patients than in controls. Furthermore, these genotypes conferred an increased mortality rate. Abnormal copy numbers of *SMN1* were subsequently shown to be a risk factor for ALS in an independent French sample[21]. Involvement of CNVs in ALS susceptibility was also demonstrated in a cytogenetic study on 85 ALS patients describing high rates of chromosomal aberrations[22].

Recent technological developments have enabled the high-throughput genotyping of large cohorts on a genome-wide scale. High-density SNP arrays can assay >300,000 SNPs throughout the genome. Several genome-wide SNP genotyping studies in ALS have been published[6,9,10,23,24], but these studies were aimed to identify associated loci using SNP genotype data and did not study the role of CNVs in disease susceptibility. However, in addition to SNP genotype data, these arrays also generate intensity data for each SNP, which enables quantification of input template DNA and indirectly, detection of deletions and duplications. Due to the high density of markers, it is possible to achieve a resolution that is >100 times higher than that of karyograms[25]. This technology has enabled a comprehensive analysis of CNVs in the human genome, and their association with disease traits.

In order to achieve a comprehensive understanding of genetic risk factors that underlie ALS susceptibility, all forms of genetic variation need to be addressed. For that reason we examined the role of copy-number variation in sporadic ALS on a genome-wide scale in a large cohort of patients and controls.

METHODS

Participants

We analysed SNP data of 461 sporadic ALS patients and 450 healthy controls, recently generated for a genome-wide association study[9,10]. All patients were diagnosed by specialised clinicians in the University Medical Centre Utrecht, which is the national referral centre for ALS in The Netherlands. Patients fulfilled criteria of probable or definite ALS according to the 1994 El Escorial criteria[26]; subjects with a positive family

history of ALS were excluded for this study. A total of 450 unrelated, healthy controls frequency-matched for age, gender and ethnicity, were recruited in the outpatient clinic of the Department of Neurology in the University Medical Centre Utrecht. Patients and controls were of self-declared Dutch descent with all four grandparents originating from The Netherlands. No immediate family relationships were observed between included samples, as indicated by the presence of more than 200,000 concordant genotypes. All study subjects included in the study gave written informed consent and the local Medical Ethics Review Boards approved all procedures.

Procedures

All DNA samples used in this study were isolated from fresh venous blood with a salting out procedure and genotyped using HumanHap300 Genotyping BeadChips (Illumina, San Diego, CA, USA). All procedures were performed according to the manufacturer's protocol. In short, 750 ng of genomic DNA was amplified, fragmented and hybridised to the array. Products were then fluorescently labelled and scanned using the Illumina Beadstation scanner. Raw data was then uploaded in Beadstudio v2·0 genotyping software (Illumina, San Diego, CA, USA) for further analysis.

HumanHap300 BeadChip data can be used for detection of CNVs by visual inspection of plotted intensity data[27]. At the time of study, there was no adequate, publicly available tool for the automated analysis of these data. Because visual inspection of our sample would be prohibitively time-consuming, we automated the process of CNV detection, by employing a script that was optimised using data of a training set of visually inspected samples (described in detail in supplementary methods).

The training set consisted of 195 controls that were analysed using methods described elsewhere[27,28]. In short, SNP data were visually inspected with the Genome Viewer tool within Beadstudio v2·0. Chromosomal regions that displayed aberrant patterns in log2 R ratio (a measure of the signal intensity at a given locus) and the B allele frequency (an estimate of the ratio between the signals of the two alleles for each SNP) were bookmarked, and used as a reference set for the optimisation of the script (figure 1). For the training set we considered aberrations of ≥10 SNPs because these have an optimal ratio of sensitivity and specificity[27]. The method discriminates between three types of CNVs: heterozygous deletions, homozygous deletions and duplications (for details see supplementary methods).

For heterozygous deletions we initially identified series of 4 homozygous SNPs with ≥3 SNPs with a negative log2 R ratio, deviating ≥2 standard deviations from the chromosomal mean. The call was then extended in forward and backward directions until the log2 R ratio deviated ≤ 1SD, or when a SNP had a heterozygous genotype call. For duplications we detected series of 4 SNPs with ≥3 SNPs having a log2 R ratio ≥2 SD above the chromosomal mean, with B allele frequency values <0·4 or. >0·6. Additionally series of 3 SNPs were detected with ≥2 SNPs having B allele frequency values of 0·2–0·4 or 0·6–0·8, without values of 0·4–0·6. The call was then extended in forward and backward directions, until none of the following conditions was met: a B allele frequency of 0·2–0·4 or 0·6–0·8 or a B allele frequency of ≥ 0·56 or ≤ 0·44 combined with a log2 R ratio of ≥1 SD above the chromosomal mean. Homozygous deletions were called when 3 consecutive SNPs showed log2 R ratios of ≤-3·5, and all adjacent SNPs with log2 R ratios of ≤-1·0. After this first data reduction step, we merged adjacent calls that were separated by ≤3 SNPs, assuming that these were in fact part of one CNV. We then discarded calls of heterozygous deletions and duplications consisting of ≤4 SNPs to reduce the false-positive rate. Next we applied more stringent criteria to the remaining CNV calls, to further reduce false-positive rate (supplementary methods).

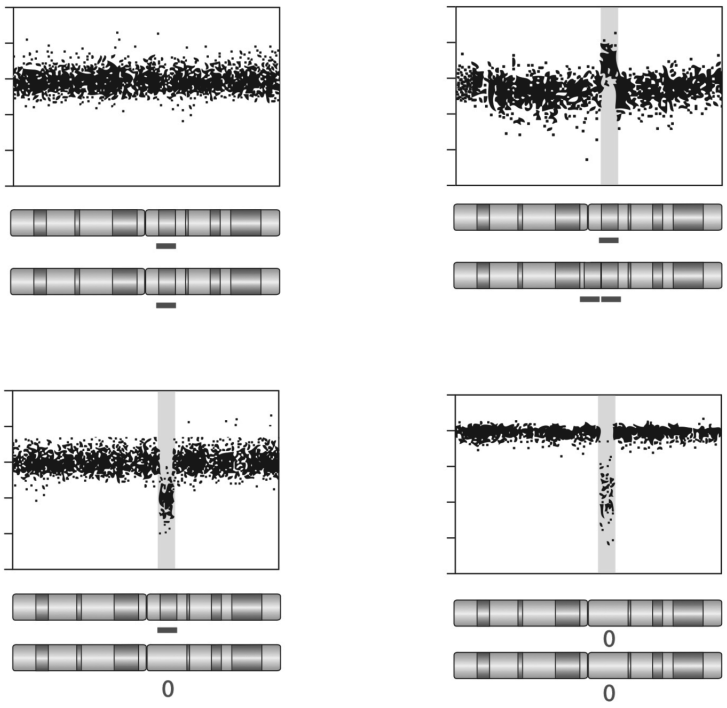


Figure 1: SNP signal intensity profiles of different CNV types.
A graphic representation of different types of CNVs and their impact on SNP intensity profiles, which can be used to detect CNVs. The signal intensity of each SNP (represented by black dots) is expressed as the log2 R ratio. The x-axis corresponds to the chromosomal positions (not to scale). In the normal (diploid) situation, the Log R ratio has a value of ~0 (A). A duplication of a chromosomal segment will cause an increase in intensity, due to the presence of 3 rather than 2 copies (B). A heterozygous deletion is marked by a negative deflection in signal intensity (C) and a homozygous deletion by an even sharper decrease in intensity (D), due to the presence of only one or no copies (please note the different scales on the y-axis).

In the training set we identified 344 CNVs, consisting of 2 homozygous deletions, 133 heterozygous deletions and 209 duplications. With the script we achieved a sensitivity of 100% for the homozygous deletions, 86% for the heterozygous deletions and 89% for the duplications. We regarded CNVs, called by the script but absent in the training set as false-positives. This was 0% for homozygous deletions, 5% for heterozygous deletions and 14% for duplications.

We applied the script on the data of 461 patients and 450 controls. We discarded samples with a log2 R ratio SD> 0·3 and samples that were identified as outliers with regard to the number of CNV calls (upper quartile + 1.5*(interquartile range)). In total 55 (11·9%) patient samples and 46 (10·2%) control samples were discarded, which was not significantly different when comparing both groups (χ^2 test, p=0·41). Because the duration of storage of DNA samples of ALS patients was longer than that of controls (median time for ALS patients was 3·2 years and 0·5 years for controls), and this could theoretically influence DNA quality, we tested for correlation between duration of storage and number of CNV calls per

individual, but found none (ρ = -0·03, p=0·41).

In order to validate our approach and to estimate the false positive rate, we selected 10 detected CNV loci for confirmation experiments using real-time quantitative PCR. We selected genomic regions that were not reported in the Database of Genomic Variants (http://projects.tcag.ca/variation, downloaded 11 April 2007) and were thus considered novel. Primers were designed with Primer Express v2·0 (Applied Biosystems, Foster City, CA, USA) (see supplementary table 1 for used primer sequences). We used primers for the albumin gene as a reference to correct for varying input amounts of DNA. To guarantee that amplicons were of unique genomic sequence we used the in-silico PCR tool of the UCSC Genome Browser (http://genome.ucsc.edu/) and examined melting curve profiles of the amplicons. In addition, we ensured that the primer sequences did not contain known SNPs present in the dbSNP database (build 125). PCR reactions were performed using the iQ SYBR Green Supermix system with a standard PCR protocol in a MyiQ Real-Time PCR cycler (Bio-Rad, Hercules, CA, USA). 10 ng of genomic DNA was used in a reaction volume of 25 µl with 12·5 µl 2x SYBRGreen Supermix and forward and reverse primers in concentrations of 400 nM. Reactions were performed in quadruplicate and each experiment was performed at least twice. Copy-numbers were calculated with methods described previously[29]. Of the loci selected for qPCR experiments, one heterozygous deletion could not be validated, while all other 9 loci were confirmed (supplementary table 1).

To further validate our results we analysed our data with PennCNV, a recently published hidden Markov model based algorithm to detect CNVs in Illumina SNP data[30]. Of the CNVs called by our script, 97·0% of heterozygous deletions and 86·6% of duplications were also detected with PennCNV. However, 59·0% of homozygous deletions were missed by PennCNV, including all qPCR-validated ones. In our training set we found a sensitivity comparable to our script: 100% of homozygous deletions, 89·2% of heterozygous deletions and 91·6% of duplications were detected. However, we found a false-positive rate of 69·9%. So although there was a high concordance between the two algorithms, PennCNV showed a much higher false-positive rate and therefore we only used CNVs as called by our script for further analyses.

We tested each CNV locus for association by performing a Fisher exact test. CNV loci were defined by grouping overlapping CNVs, and comparing these with CNV loci present in the Database of Genomic Variants. If a CNV overlapped a CNV locus reported in the Database of Genomic Variants (http://projects.tcag.ca/variation/; genome build 36, downloaded 5 October 2007), we assigned it the corresponding locus ID from the database. Otherwise we assigned a new locus ID. We performed tests for gains and losses separately.

For assessment of gene content we used gene annotation data from www.biomart.org (http://www.biomart.org/, NCBI36) and CNV start and stop positions corresponding to the 2006 genome assembly (hg18). To investigate gene functions we downloaded pathway annotations from the KEGG database (www.genome.jp/kegg/) and performed a Gene Ontology (GO) analysis with the on-line bioinformatics tool DAVID (http://david.abcc.ncifcrf.gov/home.jsp).

Statistical analysis

Statistical procedures were performed with SPSS v12·0·2 (SPSS Inc, Chicago, Ill., USA) and JMP 6·0·0 (SAS Institute Inc, Cary, NC, USA). For testing for association of CNV loci, we performed a Fisher exact test per Locus ID and used the Bonferroni correction to correct for multiple testing. For comparison of CNV size we performed a 2-tailed Mann-Whitney test, because of non-normality of the data. For comparison of group-specific genes as proportion of total detected genes, we performed a Pearson's Chi-square test (1 degree of freedom). For all tests we considered $p<0·05$ to be significant.

Role of the funding source
The study sponsors had no role in the study design, data collection, data analysis, data interpretation or writing of the report. The corresponding author had full access to all the data and had the final responsibility for the decision to submit for publication.

RESULTS

Table 1 shows the characteristics of the 810 patients and controls included in the study. In total we detected 2,328 CNVs ranging from 1·4 kb to 2·5Mb (median 68 kb) in size. The number and size of detected deletions and duplications per individual did not differ significantly in patients versus controls (table 2 and supplementary figure 1). When analysed per CNV type, duplications and heterozygous deletions were of comparable size in patients and controls. Homozygous deletions were on average larger in controls (p=0·034), but this was not significant after Bonferroni correction. The median number of genes encompassed per CNV was comparable in patients and controls (p=0·88 for duplications, p=0·55 for heterozygous deletions and p=0·67 for homozygous deletions) (supplementary figure 2).

The 2,328 CNVs that were detected in our study mapped to 935 non-overlapping loci (on average 2·49 CNVs per locus). We performed a test for association for each locus. Four loci had nominal p-values of <0·05 (table 3) and of these four, one was more prevalent in ALS patients. After Bonferroni correction for multiple testing, however, none of these remains significant.

By comparing the CNVs in our study with those previously reported in healthy controls deposited in the Database of Genomic Variants (http://projects.tcag.ca/variation), we identified 572 CNVs (24·6%) as novel. The number of novel CNVs was similar in patients (288 (25·1%)) and controls (284 (24·0%)).

Table 1: Characteristics of the study population

	Patients	**Controls**
Subjects (n)	406	404
Sex		
Male	246 (60·6%)	246 (60·9%)
Female	160 (39·4%)	158 (39·1%)
Type of onset		
Spinal	278 (68·5%)	–
Bulbar	128 (31·5%)	–
Median age at onset (years)	60 (20–86)	–
Median age at blood sampling (years)	61 (23–87)	61 (23–83)

Numbers of patients and controls with percentages in parentheses. Median age at onset and at blood sampling is given with range in parentheses. Age at onset is the date of first manifestation of weakness.

Table 2: Characteristics of detected CNVs in patients and controls

	Patients	Controls	P value
Duplications			
N	685	731	
Average per individual (range)	1·69 (0–12)	1·81 (0–10)	0·305
Median size in kilobases (range)	71·4 (5·6–2487·4)	77·9 (1·4–2014·2)	0·425
Heterozygous deletions			
N	437	431	
Average per individual (range)	1·08 (0–7)	1·07 (0–5)	0·843
Median size in kilobases (range)	61·9 (3·8–2240·6)	59·0 (4·0–1348·6)	0·766
Homozygous deletions			
N	22	22	
Average per individual (range)	0·05 (0–2)	0·05 (0–1)	0·869
Median size in kilobases (range)	13·3 (1·6–1626·9)	33·6 (1·7–190·2)	0·034

N is the total number of detected CNVs for patients and controls. The average number and the median size of detected CNVs per individual are given, with range in parentheses. P values were calculated with the Mann-Whitney U test (not corrected for multiple testing).

Table 3: CNV loci with nominal p values of <0·05

Locus	Chr	Start	End	Gain/ Loss	N Patients	N Controls	P value	RefSeq genes
NL_010	1	41119815	41147030	Gain	8	1	0·038	No genes
LC3220	16	76578045	77657555	Loss	0	6	0·015	CLEC3A; WWOX
LC3752	22	23696411	24240667	Gain	8	19	0·032	CRYBB3; CRYBB2; LOC91353; LRP5L; CRYBB2P1
LC3752	22	23696411	24240667	Loss	2	9	0·036	CRYBB3; CRYBB2; LOC91353; LRP5L; CRYBB2P1

Locus identifiers corresponding to CNV locus IDs in the Database of Genomic Variants were used if a CNV locus had previously been reported ("LC" loci). The "NL" locus does not overlap a previously reported CNV locus. Chromosomal start and end positions correspond to NCBI reference sequence build Build 36. Counts of CNVs for each locus are given for patients and controls. A Fisher exact test was performed for each CNV locus, separately for gains (i.e. duplications) and losses (i.e. heterozygous and homozygous deletions). Nominal (i.e. not corrected for multiple testing) 2-tailed p values are given. Genes that partially or completely fall within CNV loci are shown.

Table 4: Genes located in CNVs in patients and controls

	Patients	Controls	P value
Duplicated genes			
Total	883 (100%)	912 (100%)	–
Novel	304 (34%)	320 (35%)	0·769
ALS or control-specific	257 (29%)	268 (29%)	0·896
Heterozygously deleted genes			
Total	390 (100%)	323 (100%)	–
Novel	177 (45%)	69 (21%)	$1{\cdot}86\text{x}10^{-11}$
ALS or control-specific	155 (40%)	51 (16%)	$2{\cdot}15\text{x}10^{-12}$
Homozygously deleted genes			
Total	11 (100%)	4 (100%)	–
Novel	3 (27%)	1 (25%)	0·930
ALS or control-specific	1 (9%)	1 (25%)	0·423

For each category genes are divided by "total", "novel" and "ALS or control-specific". "Total" denotes the total number of genes detected for each CNV type. "Novel" genes denote genes that are located in genomic regions not deposited in the Database of Genomic Variants. "ALS or control-specific" genes were exclusively observed in either ALS patients or controls, had not been reported previously and are thus considered specific to either ALS patients or controls. In parentheses are the proportions relative to the total number of detected genes, in percentages. P values were calculated using a χ^2 test, comparing proportions between patients and controls.

We further investigated genes that were deleted or duplicated exclusively in patients (i.e. not observed in controls in our study, and not present in the Database of Genomic Variants). In total 2,021 genes were affected by the 2,328 detected CNVs. 406 genes in CNVs were exclusively observed in ALS patients: 155 genes in heterozygous deletions, 257 genes in duplications and 1 gene in a homozygous deletion; 7 genes were found in more than one CNV type. The number of genes that were duplicated exclusively in either ALS patients or controls was not significantly different (table 4). However, the number of heterozygously deleted genes that were exclusively found in ALS patients (155 (40%)) was significantly higher than the number of deleted genes exclusively observed in controls (51 (16%)) ($p=2{\cdot}15\text{x}10^{-12}$ (χ^2 test)). The 155 ALS-specific genes were observed in 54 patients; the 51 control-specific genes in 29 controls ($p=0{\cdot}004$ (χ^2 test)). The 155 ALS-specific genes in heterozygous deletions are shown in online supplementary table 2. Only one ALS-specific homozygously deleted gene was found.

To assess the functional relevance of the ALS-specific deleted genes, we examined which biochemical pathways (KEGG database, www.genome.jp/kegg) and Gene Ontology (GO) categories were represented by these genes. Of the 155 genes, 19 had a KEGG annotation, while 101 had a GO annotation. The KEGG annotated genes represented 24 biochemical pathways including oxidative phosphorylation, regulation of actin cytoskeleton and cytokine-cytokine receptor interaction (online supplementary table 2). The GO-annotated genes represented biological processes including cell cycle and mitotic checkpoint (online supplementary table 3).

DISCUSSION

In this study we performed a genome-wide search for CNVs in ALS patients and healthy controls to identify ALS-associated CNVs. CNVs are abundantly present in the human genome and represent a considerable part of genetic variation between individuals. CNVs are associated with several complex diseases and are shown to directly affect gene expression independent from SNPs. The study of CNVs appears, therefore, essential to decipher genetic susceptibility to complex disease traits. In contrast to the attention paid recently to genome-wide SNP association studies in ALS, the role of CNVs has not yet been systematically addressed.

We found a high number of rare CNVs, each present in only a small number of individuals, which is consistent with recent reports on CNVs in control populations[11,31]. The number of genes that were exclusively deleted in ALS patients was highly significantly increased compared to the number of genes that were exclusively deleted in healthy controls. Although the control-specific deleted genes may ultimately also be found in ALS patients and a number of the ALS-specific genes in controls, the significant excess of ALS-specific genes compared to controls strongly suggests that a subset of the ALS-specific-genes is indeed disease-related. As each of these genes was deleted in only one or two ALS patients this finding supports the hypothesis of rare genetic variants being contributors to ALS pathogenesis[32]. Several previous candidate gene studies in sporadic ALS reported the presence of rare pathogenic variants. For example, deletions in the tail of heavy neurofilament subunit (*NFH*) were found in 4 of 452 patients (0·88%) and missense mutations in the angiogenin (*ANG*) gene were found in 11 of 1370 patients (0·80%)[8,33]. In contrast, the recently reported associations of SNPs in *ITPR2*, *DPP6* and *FLJ10986*, were common variants (the frequency of the minor allele in patients was 11%, 42% and 39%, respectively)[6,9,10]. Hence, both common and rare genetic variants appear to be relevant to ALS[34].

To assess the functional relevance of the identified deleted genes, we determined which biological processes and biochemical pathways were represented using GO analysis and the KEGG database. These analyses identified several biochemical pathways that appear relevant to ALS pathogenesis. Interestingly, several of these pathways, including oxidative phosphorylation and cell cycle regulation showed differential expression in *SOD1* G93A mice, an animal model of ALS, as well as in autopsy specimens from sporadic ALS patients[35-37]. One of the genes involved in oxidative phosphorylation is cytochrome c oxidase subunit VIIc (*COX7C*), a component of the complex IV of the mitochondrial respiratory chain, which was heterozygously deleted in one patient. Mitochondrial dysfunction has been implicated in ALS pathogenesis previously, and a deletion in a related COX gene was found in a patient with an ALS phenotype[1,38].

Several other identified genes are biologically plausible candidates as ALS susceptibility genes, including annexin 5 (*ANXA5*), gem-associated protein 6 (*GEMIN6*) and myotubularin related protein 7 (*MTMR7*). Annexin 5 (*ANXA5*) is essential for survival and neurite growth of cortical neurons in vitro and was mentioned as functional candidate gene for ALS previously[39,40]. Gem-associated protein 6 (*GEMIN6*) is in direct interaction with the survival of motor neuron gene (*SMN*), previously shown to be associated with ALS[20,21,41]. GEMIN6 forms part of the SMN complex, which plays an essential role in the production of spliceosomal small nuclear ribonucleoproteins (snRNPs). Its activity strongly decreases when GEMIN6 is reduced in vitro[42]. *MTMR7* is expressed specifically in the brain and is thought to be involved in the phosphatidylinositol signalling pathway, recently implicated in ALS pathogenesis[9,43]. Further studies will have to reveal the true role of these deletions in ALS pathology, for example by replication in independent genome-wide datasets followed by screening large cohorts of patients and controls with targeted assays.

The major advantage of using high-density SNP arrays is the high genomic resolution that can be achieved[25]. The median size of the detected CNVs in this study was 68 kb, which is about two orders of a magnitude higher than that of traditional karyotyping that will typically detect chromosomal aberrations of >5-10 Mb[25]. The size of the detected CNVs in our study are comparable to CNVs reported by others. For example, one study reported a median size of 81 kb of CNVs detected on a SNP-based platform and 228 kb for CNVs detected with large-insert clone arrays[11]. It is therefore unlikely that an associated CNV of moderate size (50-100kb) was missed with our approach. Nevertheless, it has been shown that algorithms designed to detect CNVs on SNP-based platforms inevitably suffer from a certain false-positive and false-negative rate. The estimated false-negative reading from our current script compared to visual inspection was 14% for heterozygous deletions, and 11% for duplications. Similarly, the false-positive rate was 5% for deletions and 14% for duplications. In comparison, one of the largest SNP-based CNV study up to now, used an algorithm with an estimated false-positive rate of 6% and a false-negative rate of 18% for both duplications and deletions[11]. When the performance of that algorithm was compared to the previously identified CNVs in a well-characterized HapMap sample a false-negative rate of 24% was found. Apparently, no currently available algorithm comphrensively detects all CNVs, especially genomic variants <30kb. We therefore cannot exclude the existence of a pathogenic variant that is too small to be detected with our platform.

Despite the high genomic resolution, the coverage of the SNP array used for our study was especially targeted at unique, non-duplicated regions in the human genome and designed to optimally capture common haplotypes in the HapMap sample from European descent. Not all regions in the genome are, therefore, equally well covered. SNPs in certain regions that are prone to genomic rearrangements are technically difficult to genotype and are frequently excluded from probe sets on SNP arrays[25]. Consequently, some regions could not be assessed. This probably explains why the ALS-associated SMN locus, known to be highly variable in copy-number, was not covered by the HumanHap300 array. New-generation SNP arrays now include additional probes in these recurrent CNV regions, which form particularly interesting targets for follow-up studies.

In conclusion, we present a first exploration of genomic copy-number variation in sporadic ALS. Our data suggests that for the regions surveyed and provided the sample size of the study, no common CNVs are associated with ALS susceptibility. However, our conservative and unbiased approach to detecting CNVs in cases and controls clearly shows a significantly increased number of deleted genes specific to ALS patients representing pathways that appear relevant to ALS. Rare CNVs may play an important role in ALS pathogenesis and warrant further studies.

REFERENCES

1. Shaw PJ. Molecular and cellular pathways of neurodegeneration in motor neurone disease. *J Neurol Neurosurg Psychiatry* 2005;76:1046-57.
2. Lacomblez L, Bensimon G, Leigh PN, et al. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. *Lancet* 1996;347:1425-31.
3. Pasinelli P, Brown RH. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci* 2006;7:710-23.
4. Simpson CL, Al-Chalabi A. Amyotrophic lateral sclerosis as a complex genetic disease. *Biochim Biophys Acta* 2006;1762:973-85.
5. Graham AJ, Macdonald AM, Hawkes CH. British motor neuron disease twin study. *J Neurol Neurosurg Psychiatry* 1997;62:562-9.

6. Dunckley T, Huentelman MJ, Craig DW, et al. Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N Engl J Med* 2007;357:775-88.
7. Lambrechts D, Storkebaum E, Morimoto M, et al. VEGF is a modifier of amyotrophic lateral sclerosis in mice and humans and protects motoneurons against ischemic death. *Nat Genet* 2003;34:383-94.
8. Greenway MJ, Andersen PM, Russ C, et al. ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nat Genet* 2006;38:411-3.
9. van Es MA, Van Vught PW, Blauw HM, et al. ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet Neurol* 2007;6:869-77.
10. van Es MA, van Vught PW, Blauw HM, et al. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat Genet* 2007.
11. Redon R, Ishikawa S, Fitch KR, et al. Global variation in copy number in the human genome. *Nature* 2006;444:444-54.
12. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 2007;315:848-53.
13. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. *Nat Rev Genet* 2006;7:85-97.
14. Hollox EJ, Huffmeier U, Zeeuwen PL, et al. Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet* 2007.
15. Inoue K, Lupski JR. Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* 2002;3:199-242.
16. Fellermann K, Stange DE, Schaeffeler E, et al. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am J Hum Genet* 2006;79:439-48.
17. Aitman TJ, Dong R, Vyse TJ, et al. Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature* 2006;439:851-5.
18. Gonzalez E, Kulkarni H, Bolivar H, et al. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 2005;307:1434-40.
19. McCarroll SA, Altshuler DM. Copy-number variation and association studies of human disease. *Nat Genet* 2007;39:S37-42.
20. Veldink JH, Kalmijn S, Van der Hout AH, et al. SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. *Neurology* 2005;65:820-5.
21. Corcia P, Camu W, Halimi JM, et al. SMN1 gene, but not SMN2, is a risk factor for sporadic ALS. *Neurology* 2006;67:1147-50.
22. Meyer T, Alber B, Roemer K, et al. High rate of constitutional chromosomal rearrangements in apparently sporadic ALS. *Neurology* 2003;60:1348-50.
23. Cronin S, Berger S, Ding J, et al. A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum Mol Genet* 2007.
24. Schymick JC, Scholz SW, Fung HC, et al. Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol* 2007;6:322-8.
25. Carter NP. Methods and strategies for analyzing copy number variation using DNA microarrays. *Nat Genet* 2007;39:S16-21.
26. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J Neurol Sci* 1994;124 Suppl:96-107.
27. Peiffer DA, Le JM, Steemers FJ, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 2006;16:1136-48.
28. Simon-Sanchez J, Scholz S, Fung HC, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet* 2007;16:1-14.
29. Weksberg R, Hughes S, Moldovan L, et al. A method for accurate detection of genomic microdeletions using real-time quantitative PCR. *BMC Genomics* 2005;6:180.
30. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* 2007;17:1665-74.
31. Zogopoulos G, Ha KC, Naqib F, et al. Germ-line DNA copy number variation frequencies in a large North American

population. *Hum Genet* 2007;122:345-53.
32. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001;69:124-37.
33. Al-Chalabi A, Andersen PM, Nilsson P, et al. Deletions of the heavy neurofilament subunit tail in amyotrophic lateral sclerosis. *Hum Mol Genet* 1999;8:157-64.
34. Schymick JC, Talbot K, Traynor BJ. Genetics of sporadic amyotrophic lateral sclerosis. *Hum Mol Genet* 2007;16 Spec No. 2:R233-42.
35. Ferraiuolo L, Heath PR, Holden H, et al. Microarray analysis of the cellular pathways involved in the adaptation to and progression of motor neuron injury in the SOD1 G93A mouse model of familial ALS. *J Neurosci* 2007;27:9201-19.
36. Lederer CW, Torrisi A, Pantelidou M, et al. Pathways and genes differentially expressed in the motor cortex of patients with sporadic amyotrophic lateral sclerosis. *BMC Genomics* 2007;8:26.
37. Jiang YM, Yamamoto M, Kobayashi Y, et al. Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis. *Ann Neurol* 2005;57:236-51.
38. Comi GP, Bordoni A, Salani S, et al. Cytochrome c oxidase subunit I microdeletion in a patient with motor neuron disease. *Ann Neurol* 1998;43:110-6.
39. Han S, Zhang KH, Lu PH, et al. Effects of annexins II and V on survival of neurons and astrocytes in vitro. *Acta Pharmacol Sin* 2004;25:602-10.
40. Kasperaviciute D, Weale ME, Shianna KV, et al. Large-scale pathways-based association study in amyotrophic lateral sclerosis. *Brain* 2007;130:2292-301.
41. Veldink JH, van den Berg LH, Cobben JM, et al. Homozygous deletion of the survival motor neuron 2 gene is a prognostic factor in sporadic ALS. *Neurology* 2001;56:749-52.
42. Feng W, Gubitz AK, Wan L, et al. Gemins modulate the expression and activity of the SMN complex. *Hum Mol Genet* 2005;14:1605-11.
43. Mochizuki Y, Majerus PW. Characterization of myotubularin-related protein 7 and its binding partner, myotubularin-related protein 9. *Proc Natl Acad Sci U S A* 2003;100:9768-73.

SUPPLEMENTARY MATERIAL

Supplementary methods

195 control samples were visually inspected by an experienced data analyst as described previously[1,2] using 2 parameters computed with the Beadstudio genotyping module: the log2 R ratio and the B allele frequency.

The log2 R ratio value is a measure of the signal intensity at a given locus, compared to a reference group. It is calculated as log2 of the ratio of the observed and expected normalised intensity values (R value) at a given locus. The expected R is calculated from the intensity values of a reference group[1]. A diploid (normal copy-number) signal will have a value of ~0.

The B allele frequency gives an estimate of the proportion of B alleles at each locus. The B allele frequency of an individual SNP is computed by calculating the allelic intensity ratio of the two alleles (Θvalue), and transforming this value to a value corrected for position relative to the position of the 3 genotype clusters of the reference group[1]. An individual with an AA genotype at a given locus will thus have a B allele frequency of ~0, an AB genotype would have a value of ~0.5 and a BB genotype a value of ~1.

Abnormal copy-numbers are indicated by deviations from expected patterns of log2 R ratio and B allele frequency when plotted along the genome. For example, a heterozygous deletion (A/- or B/- genotype) will be marked by a downward deflection of the log2 R ratio, and absent heterozygous genotype calls. Alternatively, a duplication (for example AA/B or A/BB genotypes) will show an increase in log2 R ratio, and heterozygous SNPs will have B allele frequencies of around 0.33 and 0.66 (proportion of the B allele is 1/3 and 2/3, respectively), instead of ~0.5.

Based on the observations made in the training set, we formulated an algorithm that captured as much as possible of the aberrations, while limiting false positive. The probability to make a false positive CNV call depends on the number of involved markers: the false-positive rate increases when including CNVs, indicated by fewer deviating SNPs. The optimal sensitivity/specificity ratio is achieved when CNVs encompass ≥10 deviating SNPs[1]. We aimed at capturing as much of the aberrations of ≥10 SNPs, while not excluding all small (5-9 SNPs) CNV calls.

For heterozygous deletions we first identified series of 4 homozygous SNPs with ≥3 SNPs with a negative log2 R ratio, deviating ≥2 standard deviations from the chromosomal mean. The call was then extended in forward and backward directions until a log2 R ratio deviated ≤ 1SD, or when a SNP had a heterozygous genotype call.

For duplications we detected series of 4 SNPs with ≥3 SNPs with a log2 R ratio ≥2 SD above the chromosomal mean, with B allele frequency values >0.6 or <0.4. Additionally, series of 3 SNPs were detected with ≥2 SNPs having B allele frequency values of 0.2-0.4 or 0.6-0.8, without values of 0.4-0.6. The call was then extended in forward and backward directions, until none of the following conditions was met: a B allele frequency of 0.2-0.4 or 0.6-0.8 or a B allele frequency of ≥ .56 or ≤ .44 in combination with a log2 R ratio of ≥1 SD above the chromosomal mean.

Homozygous deletions were called when 3 consecutive SNPs showed log2 R ratios of ≤-3.5, and all adjacent SNPs with log2 R ratios of ≤-1.0.

We discarded samples with an SD of the log2 R ratio of ≥0.3, because these generated disproportional noise, due to irregularities in hybridization intensity signals. Likewise we discarded samples that generated a number of CNV calls that was identified as outlier (upper quartile + 1.5*(interquartile range)).

After this first data reduction step, we merged adjacent calls that were separated by ≤3 SNPs, assuming that these were in fact part of one CNV. We then discarded calls of heterozygous deletions and duplications consisting of ≤4 SNPs to reduce the false-positive rate.

Next we applied more stringent criteria to the remaining CNV calls: heterozygous deletions of ≥9 SNPs were called if ≥4 consecutive SNPs showed a log2 R ratio of ≥2SD below the chromosomal mean, or ≥3 SNPs with a log2 R ratio of ≥3 SD below the mean. Heterozygous deletions of 5-8 SNPs were called if they consisted of ≥4 consecutive SNPs with a log2 R ratio of ≥2.75 SD below chromosomal mean. Duplications were called if they consisted of ≥1 SNP with a B allele frequency of 0.2-0.4 or 0.6-0.8. Duplications that contained only homozygous genotypes were called when greater than 15 SNPs.

In the training set we identified 344 aberrations containing ≥10 SNPs. These consisted of 133 heterozygous deletions, 2 homozygous deletions and 209 duplications. With the chosen settings, we achieved a false positive rate (calls made by the algorithm that were absent in the training set) of 5% for deletions, and 14% for duplications (11% of total CNV calls). With these settings we detected 2/2 of homozygous deletions, 114/133 of heterozygous deletions (86%) and 187/209 of duplications (89%).

Of heterozygous deletions and duplications containing 5-9 SNPs, our algorithm detected 35% and 66%, respectively. Homozygous deletions showed stronger deflections in log2 R ratio than do heterozygous deletions, and are therefore require less deviating markers for a call. Of 7 homozygous deletions containing ≥3 SNPs observed in the training set, our algorithm detected 5.

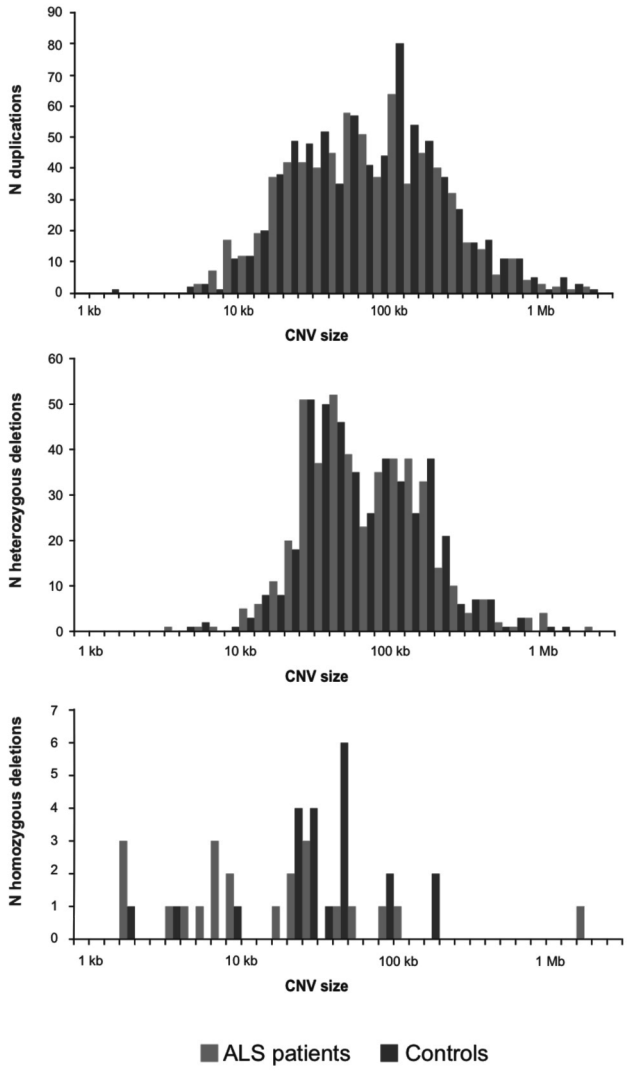
In addition to using the training set as a set of surrogate true positives, we estimated the number of false-positives in-silico, by applying the algorithm to data of 20 random samples with randomly permuted chromosomal base positions[3]. This way, all CNV calls can be expected to be false-positives. Rerunning the algorithm this way, produced 3 duplication calls and no heterozygous or homozygous deletions.

Furthermore, we selected 10 detected CNV loci for confirmation experiments using real-time quantitative PCR (see methods section). We selected genomic regions that were not reported in the Database of Genomic Variants (http://projects.tcag.ca/variation, downloaded 11 April 2007) and were thus considered novel. Two selected CNVs overlapped only partially with previously reported CNVs, but contained genes that fell outside these regions. We selected 3 loci with duplications, 4 loci with heterozygous deletions, and 3 loci with homozygous deletions. Of these loci, one heterozygous deletion could not be validated, while all other 9 loci were confirmed (supplementary table 1). Notably, 6 of the 10 CNV loci overlap with CNVs that have been reported by other authors recently (Database of Genomic Variants, 13 November 2007), which can be regarded as an extra validation.

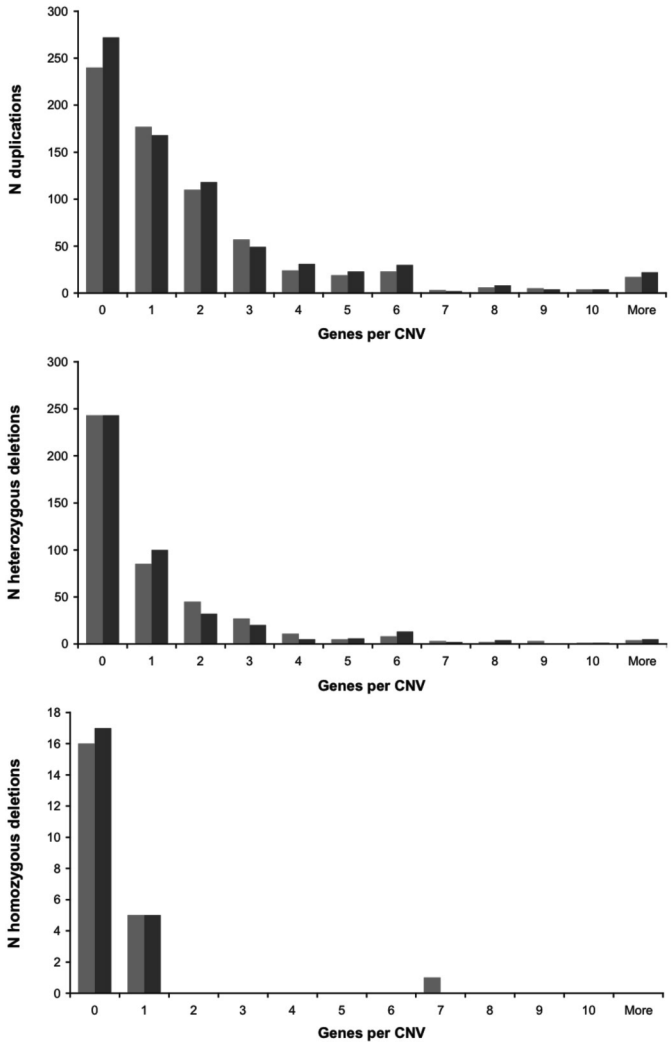
1. Peiffer DA, Le JM, Steemers FJ, et al. High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 2006;16:1136–48.
2. Simon-Sanchez J, Scholz S, Fung HC, et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. *Hum Mol Genet* 2007;16:1–14.
3. de Vries BB, Pfundt R, Leisink M, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet* 2005;77:606–1

Supplementary table 1 shows the results and used primer sequences of confirmation experiments using real-time qPCR. 9 of 10 CNV loci could be experimentally confirmed. None of the chosen regions was reported in the Database of Genomic Variants at the time of experiments.

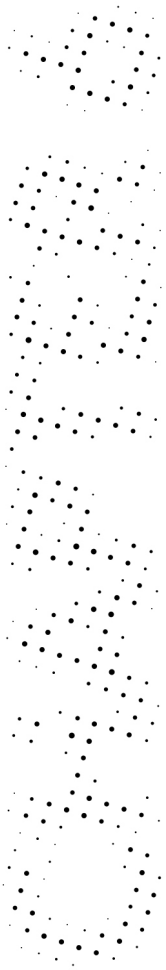
LocusID	Aberration	Chr.	Start	End	Confirmed	Forward primer	Reverse primer
NL_044	Homozygous deletion	2	49447507	49449446	Y	GGAAGCACCCGAGAAAAACTG	TAGTTCTGCTGCCACCAAACC
NL_324	Homozygous deletion	10	82869699	82875955	Y	GAATAGCAACATCTGGAACGAGTG	TCTATGCCTGTGGTGACCCTG
NL_X_001	Homozygous deletion	X	5336584	5367784	Y	AGAATAGGGTGAGTGAGGTCATGTC	CACATCTATGGGCAATGCCAC
NL_251	Heterozygous deletion	7	145599623	145892069	Y	TGCTCAGGAAGCACGCTG	AGGTATTGTCTCCCCAGAAAGATG
NL_104	Heterozygous deletion	3	128147221	128224176	N	ATTAAGGAGCGCATCCAGTCC	AACTCACCGAGTTGATGCAGC
NL_464	Heterozygous deletion	16	87729306	87821940	Y	TGGAGGCTGCAGATAGTGACC	GAGCCGTCACCCTCGTAGTC
LC0950	Heterozygous deletion	4	121050651	123291296	Y	TTTCCTTTTGAATGAGAATGCTCTATAG	CTTCATCAATTCCAGCATCAGG
NL_010	Duplication	1	41016321	41043536	Y	TTGCTTGAACCCATGTGGTG	CTCGCTCTGTAGCCCAGGC
LC2973	Duplication	16	88219799	88317399	Y	ACATAGCCCAAGGAGGCACC	GGATGTCATTGTAAATACCACCAGG
NL_080	Duplication	3	8795732	8832963	Y	GCAAGTGCCAGGAAGGACC	CTTATCTGCACAGTTAGTGGGTGG

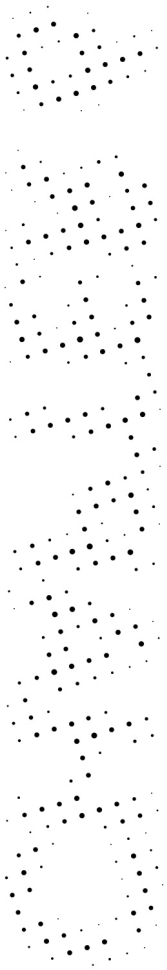


Supplementary figure 1: Size distribution of detected CNVs for patients and controls
Size is depicted on the x-axis on a logarithmic scale. The number of detected CNVs is on the y-axis (please note the different scales on the y-axis). kb=kilobase; Mb=megabase.



Supplementary figure 2: *Genic content for CNVs in ALS patients and controls*
The distribution of numbers of genes encompassed by CNVs is depicted for patients and controls. The number of genes is on the x-axis and the number of CNVs is on the y-axis (please note the different scales on the y-axis).





ANALYSIS OF GENOME-WIDE COPY NUMBER VARIATION IN IRISH AND DUTCH ALS POPULATIONS

Human Molecular Genetics (2008)

Hylke M Blauw[1,#], Simon Cronin [2,3,#], Jan H Veldink[1], Michael A van Es[1], Roel A Ophoff[4,5], Daniel G. Bradley[6], Leonard H van den Berg*[1] and Orla Hardiman*[3,7]

[1]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [2]Department of Clinical Neurological Sciences, Royal College of Surgeons in Ireland, Dublin, Ireland; [3]Department of Neurology, Beaumont Hospital, Dublin, Ireland; [4]Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [5]Neuropsychiatric Institute and Department of Human Genetics, University of California, Los Angeles, USA; [6]Smurfit Institute of Genetics, Trinity College, Dublin, Ireland; [7]Trinity College Institute of Neurosciences, Trinity College, Dublin, Ireland.

[#] Joint first authors
[*] Joint Last Authors

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is an unrelenting neurodegenerative condition characterized by adult-onset loss of motor neurons. Genetic risk factors have been implicated in ALS susceptibility. Copy number variants (CNVs) account for more inter-individual genetic variation than SNPs, and have the capacity to alter gene dose and phenotype. We sought to identify the contribution both of commonly polymorphic CNVs and rare ALS-specific CNVs to sporadic ALS. Using high density genome-wide data from 408 Irish individuals and 868 Dutch individuals and the QuantiSNP CNV-detection algorithm, we showed that no common CNV locus is significantly associated with ALS risk. However, we identified 39 recurrent CNV loci and 16 replicated ALS-specific gene dose alterations that occur exclusively in patients with ALS and do not occur in over 11,000 previously identified CNVs in the Database of Genomic Variation. Ataxin genes and the hereditary haemochromatosis locus were implicated along with *ENSG00000176605,* an uncharacterized gene on chromosome 14. Our data support the hypothesis that multiple rare CNVs may contribute risk for sporadic ALS. Future work should seek to profile the contribution of CNVs located in regions not covered on the present SNP platforms.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disorder characterized by progressive weakness of the limbs and bulbar muscles. The peak age of presentation is in the fifth to seventh decades, and average survival from symptom onset is 3-5 years. At present the only available therapy is riluzole, which increases survival by approximately 3 to 6 months. (1)

Between 2 and 10% of ALS cases are associated with familial, usually autosomal dominant, inheritance. For ~20% of familial ALS, mutations have been identified in genes, including *Cu/Zn superoxide-dismutase, dynactin 1, alsin, vesicle-associated protein B, senataxin, angiogenin* and *TAR-DNA (TDP-43).* (2-9) The cause of sporadic ALS (SALS) is less well understood, although it is widely believed that genetic risk factors are involved. (10) Indeed, over the past year genome-wide association studies for SALS have reported initial SNP associations with *FLJ10986, ITPR2* and *DPP6* (11-14), and it is likely that a clear picture of the contribution of SNPs to ALS will be uncovered as sample numbers worldwide increase.

In addition to the contribution of SNPs to inter-individual variation across the genome, it is now recognized that insertions or deletions of DNA sequence, typically ranging from 1kb to several mega-bases in size, also contribute to genetic variation and disease susceptibility. (15) These are termed copy number variations (CNVs) and comprise between 4 and 25Mb of an individual's genome. (16-17) Thus CNVs account for more nucleotide variation between two individuals than SNPs, which account for only 2.5Mb of the genome. (17) As with SNPs, individual CNV loci may be commonly polymorphic within the population, or may occur as inherited or de novo rare mutations (15-17). Loss (deletion) or gain (duplication) of genomic sequence can influence gene expression levels (18) or lead to truncated proteins with altered functionality, providing mechanisms by which CNVs may influence disease susceptibility. There are already many known CNV phenotypes - some of these exhibit Mendelian inheritance (for example, duplication of peripheral myelin protein 22 in autosomal dominant Charcot-Marie Tooth neuropathy

type 1A) (19) whereas others are sporadic and result from de novo CNV mutation (as for example, in Williams syndrome). (20) Our previous observation that CNV of the *survival of motor neuron (SMN)* genes influences susceptibility to and severity of ALS is a precedent for a role of CNVs in ALS pathogenesis. (21)

The recent development of robust, high-throughput genotyping arrays provides the opportunity to rapidly screen up to one million SNPs across the human genome. Genotyping on these arrays has been used in sporadic ALS (SALS) to examine allelic association at SNPs. (11-14) However, in addition to genotyping calls, the arrays provide a binding intensity measure at each SNP. (22) Low intensity of several contiguous SNPs suggests the presence of a deletion in that region, while high intensity at several contiguous SNPs suggests duplication. Together with genotype, these data can be used as an indirect measure to detect the positions of CNVs along each chromosome. This method will identify not only common CNV polymorphisms, but also rare and potentially de novo CNV mutations even when present in a single study participant. (22)

We have recently reported a genome-wide screen of CNVs in a Dutch sporadic ALS dataset, in which we found an excess of hemizygous gene deletions specific to ALS. (23) In the present work, we adopted a collaborative approach to further characterize the contribution of CNVs to ALS by using genome-wide SNP intensity data from the Irish and Dutch ALS populations and the well validated CNV detection algorithm QuantiSNP which can robustly detect even small CNVs. (24) We examined both association of common CNVs with ALS susceptibility and made an effort to identify rare ALS-specific genomic variants.

RESULTS

We genotyped 432 unique Irish individuals using the Illumina HumanHap 550 Beadchip (Illumina Inc., San Diego) and 911 unique Dutch individuals using the Illumina HumanHap 300 Beadchip. Genotyping methods and quality control criteria have been described previously. (12,14) B Allele Frequency (BAF) and Log R Ratio (LRR) for each SNP was recorded using BeadStudio version 3.0 (Illumina Inc) and the QuantiSNP algorithm (24) was utilized to make CNV calls. After applying the quality control criteria described in Materials and Methods, genome-wide CNV data of 408 Irish individuals (comprising 206 patients with sporadic ALS and 202 control subjects) and 868 individuals from The Netherlands (comprising 445 patients with sporadic ALS and 423 control subjects) were included in the final analysis. Table 1 shows the characteristics of the study participants.

In total, 5017 CNVs were called using 550K array data in the 408 Irish participants (on average 12.3 CNVs per genome) and 4103 CNVs were called using 300K array data in the 868 Dutch participants (on average 4.7 CNVs per genome). Table 2 shows, for each CNV type, the mean number of calls per individual and the median CNV size. As was expected, the algorithm detected more CNVs per individual using the 550K data as compared to the 300K data, due to the higher marker density on the 550K platform. However, the number of CNV calls per individual did not differ between patients and controls within the Irish or Dutch study group. Interestingly, the median length of hemizygous deletions was greater in patients with ALS compared to controls in both populations. In contrast, the length of duplications and homozygous deletions did not differ between groups.

To further assess the role of CNVs in ALS, we examined pooled CNV data from our populations, and examined both association of common CNVs and the occurrence of novel CNVs. Because breakpoints of CNVs at the same locus may vary between individuals, we examined gain and loss of copy number at each involved SNP using

pooled data from the two study populations, and tested for association with ALS. We detected 26 SNP copy number gains, corresponding to 5 chromosomal regions (Supplementary table 1) and 58 SNP copy number losses, corresponding to 16 chromosomal regions (Supplementary table 2) which showed nominal association with ALS at p values below 0.05 in the pooled dataset. However, after Bonferroni correction, none of these remained significantly associated.

To identify novel CNVs in our combined dataset, we compared our data with the Database of Genomic Variants (DGV), available at http://projects.tcag.ca/variation, which contains more than 11,000 CNVs detected in more than 2,500 individuals. (15) To avoid inclusion of false positives, we searched for recurrent novel CNVs, i.e. overlapping CNVs that were present in at least two individuals. We detected 89 recurrent CNVs mapping to 39 novel genomic loci that were exclusive to ALS patients (i.e. at CNV loci that had not been previously reported in the DGV nor in the controls in our study) (online Supplementary table 3). In 14 of these 39 recurrent CNV loci (36%) a CNV was detected in both Irish and Dutch ALS patients.

Finally, we investigated genes that were deleted or duplicated exclusively in ALS patients (i.e., not altered in any controls in our populations and not reported in the DGV). We corroborated our previously reported finding that there was an excess of ALS-specific deleted genes in the Dutch population, compared to controls: 23% of deleted genes in ALS patients were exclusive to ALS patients, while only 14% of deleted genes observed in controls were exclusive to controls ($p=5.4 \times 10^{-5}$, X^2 test). We did not observe this in the Irish population: the proportion of deleted genes that was exclusive to either ALS patients or controls was 22% in both Irish ALS patients and controls. We identified 16 genes bearing loss or gain of copy number exclusively among ALS patients, replicated in both the Irish and Dutch analyses (Table 3). Among these, there was one uncharacterized gene on chromosome 14 *(ENSG00000176605)* where a hemizygous deletion was observed in 6 individual patients. Two neighbouring genes on chromosome 8 showed copy loss in five ALS patients and copy gain in one patient. The remaining ALS-specific genes showed deletions or duplications in two or three patients.

Table 1 Characteristics the Irish and The Netherlands study populations

	Total	Male n (%)	Female n (%)	Age at onset (median, range)	Spinal onset n (%)	Bulbar onset n (%)
Ireland						
Patients with sporadic ALS	206	112 (54%)	94 (46%)	62 (23-90)	147 (71%)	59 (29%)
Controls	202	108 (53%)	94 (47%)	59 (25-84)	-	-
The Netherlands						
Patients with sporadic ALS	445	262 (59%)	183 (41%)	62 (23-87)	306 (69%)	139 (31%)
Controls	423	249 (59%)	174 (41%)	61 (23-87)	-	-

Table 2 Characteristics of detected CNVs

	IRELAND – 550K array data			THE NETHERLANDS – 300K array data		
	ALS	**Control**	**p value***	**ALS**	**Control**	**p value***
Duplications						
Total	689	726		891	883	
Mean per individual (range)	3.3 (0-9)	3.6 (0-9)	0.171	2.0 (0-6)	2.2 (0-6)	0.428
Median size, kb (range)	76.4 (0.1-3261.7)	71.8 (0.1-1578.9)	0.602	93.4 (0.2-2534.6)	98.7 (0-2914.7)	0.527
Heterozygous deletions						
Total	1549	1699		1177	1079	
Mean per individual (range)	7.6 (0-16)	8.4 (1-18)	0.006	2.71 (0-8)	2.63 (0-8)	0.467
Median size, kb (range)	21.9 (0.4-214.8)	17.4 (0.2-214.8)	<0.0001	44.1 (0.6-2240.6)	37.6 (0.6-1448.1)	0.011
Homozygous deletions						
Total	188	166		44	29	
Mean per individual (range)	0.91 (0-7)	0.82 (0-5)	0.59	0.10 (0-2)	0.07 (0-1)	0.242
Median size, kb (range)	8.8 (0.1-220.1)	5.99 (0.1-606.1)	0.609	12.5 (0.2-1626.9)	24.1 (0-190.3)	0.632

* Two-tailed Mann-Whitney U test

DISCUSSION

We conducted a genome-wide search for ALS-associated and ALS-specific CNVs in 651 patients with sporadic ALS and 625 control subjects drawn from the Irish and Dutch populations. We identified CNV loci exhibiting recurrent CNV uniquely in patients with ALS and identified genes with replicated ALS-specific dose alterations between our populations.

The recent advent of genome-wide approaches, facilitating characterization of the genome at much higher resolution than hitherto, has revealed the presence of a rich mosaic of sub-microscopic chromosomal alterations, including segments that are deleted, duplicated, inverted in orientation, inserted and translocated. (15-18,25) The influence of these genetic variants on human disease remains understudied, and it has been suggested that many smaller variations remain to be discovered. (26-28) To follow our initial characterization of CNVs in ALS (23), here we utilized QuantiSNP, a validated and sensitive algorithm for CNV detection using SNP array data, and describe the potential contribution of CNVs to ALS pathogenesis at higher resolution than hitherto and in two independent populations.

We took two approaches to uncover rare CNV mutations of potential relevance to ALS. The first, based upon overlapping CNVs, implicated 39 genomic loci as novel, recurrent and ALS-specific. Several of the genes affected by these variations are

Table 3 Genes found to have ALS-specific alteration of copy number, in both the Irish and Dutch populations

				IRELAND		THE NETHERLANDS		
Ensembl Gene ID*	Gene	Putative functiβon	Chr	ALS patients (N)	Gain/ loss	ALS patients (N)	Gain/ loss	Total
ENSG00000176605	C14orf177	Uncharacterized	14	1	Loss	5	Loss	6
ENSG00000104518	GSDMDC1	Tyrosine phosphorylation	8	5	Gain and loss	1	Loss	6
ENSG00000173201		Uncharacterized	8	5	Gain and loss	1	Loss	6
ENSG00000101846	STS	Arylsulphatase	X	1	Gain	2	Gain and loss	3
ENSG00000130021	HDHD1A	Uncharacterized	X	1	Gain	2	Gain and loss	3
ENSG00000134538	SLCO1B1	Xenobiotic and drug detoxification solute carrier	12	2	Loss	1	Gain	3
ENSG00000077713	SLC25A43	Mitochondrial solute carrier widely expressed in CNS	X	1	Loss	1	Loss	2
ENSG00000123454	DBH	Converts dopamine to noradrenalin in nerve cells	9	1	Loss	1	Loss	2
ENSG00000151963		Uncharacterized	12	1	Gain	1	Gain	2
ENSG00000156110	ADK	Catalyzes phosphate transfer to adenosine	10	1	Loss	1	Loss	2
ENSG00000174373	GARNL1	Variants previously implicated in microcephaly and Fahr disease	14	1	Loss	1	Loss	2
ENSG00000186440	OR6P1	Olfactory receptor gene family member	1	1	Gain	1	Gain	2
ENSG00000201674		Uncharacterized	X	1	Loss	1	Gain	2
ENSG00000207220		Uncharacterized	X	1	Loss	1	Gain	2
ENSG00000208843		Uncharacterized	X	1	Loss	1	Gain	2
ENSG00000212611		Uncharacterized	X	1	Loss	1	Gain	2

* These 16 genes showed copy number variation only in ALS patients and were observed in both the Irish and Dutch ALS populations. Copy number variation involving these genes was not observed in the 625 Irish and Dutch controls nor among >2,500 controls reported in the Database of Genomic Variants.

biologically plausible candidates in ALS susceptibility, including *ataxin 1 (ATXN1), ataxin 3-like protein (ATXN3L)* and the *hereditary haemochromatosis gene (HFE).* For example, variations in *HFE* significantly increased the risk of SALS in a number of different populations. (29) Furthermore, we particularly highlight deletion of *ataxin 1 (ATXN1)* in two Irish SALS patients and a duplication of the related gene *ataxin 3-like protein (ATXN3L)* in a Dutch SALS patient. Occurrence of an expanded polyglutamine repeat in *ATXN1* is the cause of autosomal dominant spinocerebellar ataxia type 1 (SCA1) (30), a neurological disease with phenotypic similarities to ALS. Over-expression

of *polyglutamine binding protein 1* (*PQBP-1)*, which interacts with *ATXN1*, induces a progressive motor-neuron disease phenotype in mice. (31) Furthermore, we have recently noted strong association of another ataxin variant in our SALS populations. (12,14) rs1551960, an intronic marker in *ataxin-2 binding protein 1 (A2BP1)*, showed an allelic association in both the Irish genome-wide SNP study (uncorrected p value 4.7×10^{-5}) and the Dutch genome-wide SNP study (uncorrected p value 0.007). The combined p value for this SNP in the two populations is 8.4×10^{-6}. A2BP1 is a trans-Golgi network protein which binds the C-terminus of ataxin 2, and modifies the pathology of spinocerebellar ataxia type 2. (32-33) Taken together, our findings raise the intriguing possibility that the ataxins and functionally related genes may influence susceptibility for ALS.

Our second approach to identify rare variants focused on gene dose alterations that were exlusively observed in ALS patients. We previously reported an excess of ALS-specific deleted genes in the Dutch population. This finding was corroborated when reanalysing our data with QuantiSNP, making a spurious finding due to CNV detection methodology unlikely. However, in the Irish population, we did not observe a similar finding. This could be explained by sampling issues due to the relatively low number of patients, or by genetic heterogeneity between the two populations. Our previously reported finding of extended linkage disequilibrium and increased genetic homogeneity in the Irish population could support the latter. (14) Because the finding of a rare variant in a disease population can be explained by chance, here we identified genes that showed a gene dose alteration exclusively in ALS patients replicating in both Irish and Dutch populations. Moreover, deletion CNVs were larger in both populations in ALS patients compared to controls, suggesting that deletions in ALS are more detrimental. Of the 16 gene alterations identified as replicate, seven have an identical effect on gene dose in each population. Among these, the most compelling result, being present in 6 ALS patients, was hemizygous loss of *ENSG00000176605*, a gene of unknown expression and function.

Common CNV polymorphisms seem not to alter risk for developing ALS. However, as the present generation of genotyping arrays do not include regions which are technically difficult to genotype or prone to genomic rearrangements, there is a bias in coverage away from areas with the highest CNV polymorphism rate. (22) For example, the previously reported *SMN* genes are not covered by the array. For this reason, we have focused on the discovery of rare and potentially pathogenic CNVs. Although it is not unlikely that these ALS-specific CNVs may ultimately be found also in controls, we reported putative mutations only where an observation was made in two or more individuals with ALS and we have further censored our data for CNVs reported in the Database of Genomic Variants. As ALS is a late-onset disease, we cannot be certain whether use of the DGV in this way might have excluded ALS related CNVs which were ascertained in younger individuals prior to disease onset. Similarly, the late onset of ALS precludes genotyping of the parents of affected individuals to differentiate inherited versus de novo CNVs as has been accomplished, for example, in CNV studies on autism. (35)

Genomic structural variants of all sizes and types may act as substrates for disease risk. (26) Reliance on commercial SNP arrays for detection of these variations thus has limitations which must be considered when interpreting our findings. First, it is recognized that the lack of "breakpoint resolution" on these arrays mean that detected CNVs may include more genomic sequence than between the first and final involved SNP. (26) This precludes delineation of the true extent of any given CNV and hence valid allele frequency counts. Additionally, the array data cannot detect the potential influence of balanced rearrangements, such as translocations and segment inversions. Finally, owing to the low density of Y chromosome markers, no CNV calls are made for the Y chromosome.

In conclusion, we further characterized the contribution of genomic copy-number variations in sporadic ALS in the Irish and Dutch populations. Using the presently available sample size and genotyping platforms, no common CNVs appear associated with SALS risk. Our data support our previous finding that rare CNVs may have an important role in ALS pathogenesis. Future studies will be necessary to confirm these observations in additional populations and to explore those genomic structural variants, which by their location or nature, have not yet been addressed in ALS research.

MATERIALS AND METHODS

Participants

The final Irish study population comprised 206 ALS patients and 202 controls; the final Dutch population comprised 445 ALS patients and 423 controls (Table 1). The Irish DNA samples were collected at the Beaumont Hospital in Dublin, Ireland and the Dutch DNA samples were collected at the ALS centre in the University Hospital in Utrecht, The Netherlands. In both Ireland and The Netherlands, all patients fulfilled the 1994 El Escorial criteria (35) for probable or definite sporadic ALS and were phenotyped by physicians with expertise in ALS. Patients with a family history of ALS have been excluded from the study. For quality control, participants with El Escorial "possible ALS" and those with atypical clinical features are not included. At both centres, patients are followed prospectively and any with revision to an alternative diagnosis removed. However, no patients genotyped in the present experiments were removed for these reasons. Control DNA samples were collected from healthy, unrelated neurologically normal individuals, either spouses of ALS patients or those accompanying non-ALS patients. Controls were frequency-matched for age, gender and ethnicity. In the Irish population all participants were of self-declared Irish-Caucasian ethnicity for at least three generations. The Dutch participants were of self-declared Dutch descent with all four grandparents originating from The Netherlands. All participants gave written informed consent and local ethics review boards approved all procedures.

Procedures

Genomic DNA was isolated from fresh venous blood using standard procedures and used in this form for both the Irish and Dutch experiments. No samples had been immortalized in cell lines prior to assay. The Irish samples were genotyped using Illumina HumanHap 550 Beadchips, which assay 561,466 SNPs. The Dutch samples were genotyped using Illumina HumanHap 300 Beadchips, which assay 317,503 SNPs. All experiments were performed according to manufacturer's protocol. Stringent quality control has been applied to the samples in our study as previously described (12,14): in particular samples with call rates below 97.5% (Irish population) or 95% (Dutch population) and samples that showed cryptic relatedness had been excluded.

CNV detection

Putative CNVs were identified using the QuantiSNP algorithm. (24) In brief, QuantiSNP is a Hidden-Markov Model-based algorithm for high-resolution CNV detection that uses Log2 R Ratio (LRR) values (a normalised measure of the signal intensity for each SNP) and B Allele Frequency (BAF) values (the frequency of the minor allele for each SNP) to generate CNV calls. As a measure of confidence, a Bayes factor is computed for each CNV. A correction for local differences in GC content is implemented in the algorithm to correct for irregularities in signal intensity.

We computed LRR and BAF values with the Beadstudio v3.0 genotyping module

(Illumina Inc.) and analysed these with QuantiSNP using default settings. We discarded all CNV calls with a Bayes factor value below 10. (24) In order to minimise the number of false-positive calls, we also excluded samples that fulfilled one of the following criteria: samples with a standard deviation of the LRR > 0.3 after GC correction; samples with a standard deviation of the BAF > 0.15 after GC correction; and samples that were outliers in terms of number of CNV calls (expressed as upper quartile + 1.5*(interquartile range)). In total, 26 samples from the Irish study and 43 samples from the Dutch study failed QC criteria. The proportion of excluded cases vs controls was not significant in Irish and Dutch populations (p=0.25 in the Irish population, p=0.07 in the Dutch population, ^{2} test). Finally, we excluded CNV calls that spanned the centromere, because the low density of probes in these regions leads to inaccurately called CNVs.

We determined false-positive rates and false-negative rates using two methods. First, we performed replicate experiments on five samples that were genotyped twice and analysed the data using QuantiSNP. In three samples genotyped on the 300K platform, 18 CNVs were detected with a Bayes factor above 10. In the replicates, 16 out of 18 CNVs (89%) were called with a Bayes factor above 10, one (5%) CNV was called with a Bayes factor below 10 and one CNV (5%) was not detected in the replicate experiment. In two samples genotyped on the 550K platform, 15 CNVs were called with Bayes factor above 10. In the two replicates, 13 (87%) of these were called with a Bayes factor above 10, one (6%) was called with a Bayes factor below 10 and one (6%) was not detected in the replicate experiment.

As a second method to determine the performance of the algorithm, we applied the algorithm on datasets of simulated CNVs. For this purpose, we constructed artificial CNVs from LRR and BAF values from three large, qPCR validated CNVs: a one-copy duplication, a one-copy (hemizygous) deletion and a two-copy (homozygous) deletion. We constructed chromosomes of randomly shuffled diploid SNPs and inserted series varying from 2 to 10 SNPs from the biologically confirmed CNVs. This way we simulated 500 CNVs per CNV length for each CNV type. We applied QuantiSNP on the simulated data and determined the false-positive and false-negative counts. At the Bayes factor cut-off value of 10, QuantiSNP detected zero false-positive CNVs per 860,085 SNPs (45 simulated chromosomes of 19,113 SNPs). The false-negative readings for each CNV length are depicted in Figure 1.

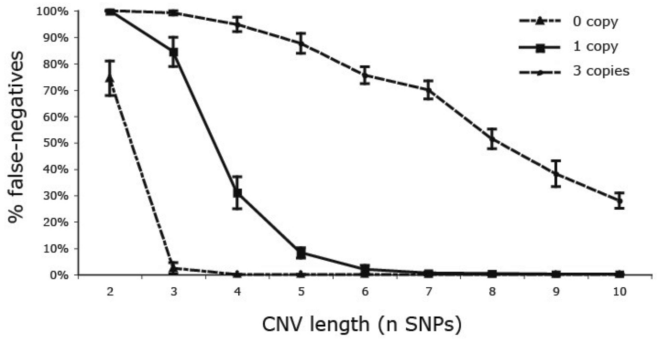


Figure 1. Performance of the CNV detection algorithm QuantiSNP. CNVs of different lengths were simulated by inserting LRR and BAF values from validated CNVs into artifical chromosomes of randomly shuffled diploid SNPs. Homozygous deletions (0 copies), hemizygous deletions (1 copy) and hemizygous duplications (3 copies) were modelled. For each CNV length 5 chromosomes containing 100 CNVs were constructed. Data points are the mean percentage of detected CNVs +/- standard deviation of 5 chromosomes.

Statistical analysis
All statistical analyses were computed using SPSS 16.0, JMP v6.0 (SAS inc.) and R. For comparisons of CNV lengths and number of CNVs per individual we used the Mann-Whitney U test, because of non-normality of the data. We ran the program STRUCTURE version 2.0 (36) using 4,118 pseudo-randomly selected unlinked markers and identified stucture differences between our populations. We did not detect population stratification, which could lead to spurious associations, between the pooled case and pooled control groups. To identify associated CNV regions, we performed the exact Cochran-Mantel-Haenszel chi-squared test per SNP using pooled Irish and Dutch data and Fisher's exact test for separate populations. To correct for multiple testing, we used the Bonferroni correction. For the rare variation analysis, CNVs were defined as recurrent when they overlapped in at least two individuals.

Role of the funding source
The study sponsors had no role in the design of the study, analysis and interpretation of the data.

Acknowledgements
We sincerely thank the patients, their relatives and referring neurologists for their participation. This research was funded by the Muscular Dystrophy Association (United States) (SC, DGB, OH), by the Health Research Board of Ireland (OH), by the Irish Institute of Clinical Neuroscience Travel Award (SC), by the Irish Motor Neurone Disease Research Foundation; by the Prinses Beatrix Fonds, the Kersten Foundation, the Netherlands Organisation for Scientific Research (NWO), the Adessium Foundation and VSB Fonds (LHvdB); and by The Brain Foundation of The Netherlands (JHV).

Conflicts of interest
None declared

REFERENCES

1 Lacomblez, L., Bensimon, G., Leigh, P.N., Guillet, P., and Meininger, V. (1996) Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/Riluzole Study Group II. Lancet, 347, 1425–1431.

2 Pasinelli, P. and Brown, R.H. (2006) Molecular biology of amyotrophic lateral sclerosis: insights from genetics. Nat. Rev. Neurosci., 7, 710-723.

3 Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X., et al. (1993) Mutations in Cu/Znsuperoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature, 362, 59–62.

4 Puls, I., Jonnakuty, C., LaMonte B.H., Holzbaur, E.L., Tokito, M., Mann, E., Floeter, M.K., Bidus, K., Drayna, D., Oh, S.J., et al. (2003) Mutant dynactin in motorneuron disease. Nat. Genet., 33, 455–456.

5 Yang, Y., Hentati, A., Deng, H.X., Dabbagh, O., Sasaki, T., Hirano, M., Hung, W.Y., Ouahchi, K., Yan, J., Azim, A.C., et al. (2001) The gene encoding alsin, a protein with three guanine-nucleotide exchange factor domains, is mutated in a form of recessive amyotrophic lateral sclerosis. Nat. Genet., 29, 160-165.

6 Nishimura, A.L., Mitne-Neto, M., Silva, H.C., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J.R., Gillingwater, T., Webb, J. et al (2004) A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. Am. J. Hum. Genet., 75, 822–831.

7 Chen, Y.Z., Bennett, C.L., Huynh, H.M., Blair, I.P., Puls, I., Irobi, J., Dierick, I., Abel, A., Kennerson, M.L., Rabin, B.A., et al. (2004) DNA/RNA helicase gene mutations in a form of juvenile amyotrophic lateral sclerosis(ALS4). Am. J. Hum. Genet., 74, 1128–1135.

8 Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., et al. (2006) ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. Nat. Genet., 38, 411–413.

9 Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., et al. (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science, 319, 1668-1672.

10 Schymick, J.C., Talbot, K., and Traynor, B.J. (2007) Genetics of sporadic amyotrophic lateral sclerosis. Hum. Mol. Genet., 16, R233-R242.

11 Dunckley, T., Huentelman, M.J., Craig, D.W., Pearson, J.V., Szelinger, S., Joshipura, K., Halperin, R.F., Stamper, C., Jensen, K.R., Letizia, D., et al. (2007) Whole-genome analysis of sporadic amyotrophic lateral sclerosis. N. Engl. J. Med., 357, 775–788

12 Van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Andersen, P.M., Van Den Bosch, L., de Jong, S.W., van 't Slot, R., et al. (2007) ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. Lancet Neurol., 6, 869–877.

13 Van Es, M.A., van Vught, P.W., Blauw, H., Franke, L., Saris, C.G., Van Den Bosch, L., de Jong, S.W., de Jong, V., Baas, F., van't Slot, R.. et al., (2008) Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. Nat. Genet., 40, 29-31.

14 Cronin, S., Berger, S., Ding, J., Schymick, J.C., Washecka, N., Hernandez, D.G., Greenway, M.J., Bradley, D.G., Traynor, B.J. and Hardiman, O. (2008) A genome-wide association study of sporadic ALS in a homogenous Irish population. Hum. Mol. Genet., 17, 768-774.

15 Iafrate, A.J., Feuk, L., Rivera, M.N., Listewnik, M.L., Donahoe, P.K., Qi, Y., Scherer, S.W. and Lee, C. (2004) Detection of large-scale variation in the human genome. Nat. Genet., 36, 949-951.

16 Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W., et al. (2006) Global variation in copy number in the human genome. Nature, 444, 444-454.

17 Sebat, J. (2007) Major changes in our DNA lead to major changes in our thinking. Nat. Genet., 39, S3-S5.

18 Stranger, B.E., Forrest, M.S., Dunning, M., Ingle, C.E., Beazley, C., Thorne, N., Redon, R., Bird, C.P., de Grassi, A., Lee, C., et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science, 315, 848-853.

19 Lupski, .JR., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., et al. (1991) DNA duplication associated with Charcot-Marie-Tooth disease type 1A. Cell, 66, 219-232.

20 Ewart, A.K., Morris, C.A., Atkinson, D., Jin, W., Sternes, K., Spallone, P., Stock, A.D., Leppert, M. and Keating, M.T. (1993) Hemizygosity at the elastin locus in a developmental disorder, Williams syndrome. Nat. Genet., 5, 11-16.
21 Veldink, .JH., van den Berg, L.H., Cobben, J.M., Stulp, R.P., De Jong, J.M., Vogels, O.J., Baas, F., Wokke, J.H. and Scheffer, H. (2001) Homozygous deletion of the survival motor neuron 2 gene is a prognostic factor in sporadic ALS. Neurology 56, 749-752.
22 Carter, N.P. (2007) Methods and strategies for analyzing copy number variation using DNA microarrays. Nat. Genet., 39, S16-S21.
23 Blauw, H.M., Veldink, J.H., van Es, M.A., van Vught, P.W., Saris, C.G., van der Zwaag, B., Franke, L., Burbach, J.P., Wokke, J.H., Ophoff, R.A., et al. (2008) Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-widescreen. Lancet Neurol., 7, 319-326.
24 Colella, S., Yau, C., Taylor, J.M., Mirza, G., Butler, H., Clouston, P., Bassett, A.S., Seller, A., Holmes, C.C. and Ragoussis, J. (2007) QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 35, 2013-2025.
25 Scherer, S.W., Lee, C., Birney, E., Altshuler, D.M., Eichler, E.E., Carter, N.P., Hurles, M.E. and Feuk, L. (2007) Challenges and standards in integrating surveys of structural variation. Nat. Genet., 39, S7-S15.
26 Conrad, D.F. and Hurles, M.E. (2007) The population genetics of structural variation. Nat. Genet., 39, S30-S36.
27 de Smith, A.J., Tsalenko, A., Sampas, N., Scheffer, A., Yamada, N.A., Tsang, P., Ben-Dor, A., Yakhini, Z., Ellis, R.J., Bruhn, L., et al.(2007) Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. Hum. Mol. Genet., 16, 2783-2794.
28 Kidd, J.M., Cooper, G.M., Donahue, W.F., Hayden, H.S., Sampas, N., Graves, T., Hansen, N., Teague, B., Alkan, C., Antonacci, F., et al. (2008) Mapping and sequencing of structural variation from eight human genomes. Nature, 453, 56-64.
29 Goodall, E.F., Greenway, M.J., van Marion, I., Carroll, C.B., Hardiman, O. and Morrison, K.E. (2005) Association of the H63D polymorphism in the hemochromatosis gene with sporadic ALS. Neurology, 65, 934-937.
30 Banfi, S., Servadio, A., Chung, M.Y., Kwiatkowski, T.J. Jr., McCall, A.E., Duvick, L,A,, Shen, Y., Roth, E.J., Orr, H.T. and Zoghbi, H.Y. (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. Nat. Genet., 7, 513-520.
31 Okuda, T., Hattori, H., Takeuchi, S., Shimizu, J., Ueda, H., Palvimo, J.J., Kanazawa, I., Kawano, H., Nakagawa, M. and Okazawa, H. (2003) PQBP-1 transgenic mice show a late-onset motor neuron disease-like phenotype. Hum. Mol. Genet., 12, 711-725.
32 Shibata, H., Huynh, D.P. and Pulst, S.M. (2000) A novel protein with RNA-binding motifs interacts with ataxin-2. Hum. Mol. Genet. 9, 1303-1313.
33 Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.F., Fisk, C.J., Li, N., Smolyar, A., Hill, D.E., et al. (2006) A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell, 125, 801-814.
34 Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., et al. (2007)Strong association of de novo copy number mutations with autism. Science, 316, 445-449.
35 Brooks, B.R., Miller, R.G., Swash, M., Munsat, T.L., World Federation of Neurology Research Group on Motor Neuron Diseases. (2000) El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph. Lateral Scler. Other Motor Neuron Disord., 1, 293-299.
36 Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of populationstructure using multilocus genotype data. Genetics, 155, 945–959.

SUPPLEMENTARY MATERIAL

Supplementary Table 1 SNPs showing association with ALS by gain of copy number in the combined Irish and Dutch study populations, at uncorrected p values below 0.05

SNP	Chr.	Position	Gene	Position relative to gene	IRELAND			THE NETHERLANDS			POOLED		
					ALS (n)	Controls (n)	P	ALS (n)	Controls (n)	P	ALS (n)	Controls (n)	P
rs329747	19	58230317	FLJ32214	flanking 3'UTR	1	10	0.005	0	0	-	1	10	0.005
rs6509718	19	58240762	ZNF160	flanking 3'UTR	1	10	0.005	0	0	-	1	10	0.005
rs10501323	11	48461172	OR4A47	flanking 5'UTR	0	0	-	0	5	0.063	0	5	0.027
rs10838881	11	48343851	OR4C3	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs12282928	11	48288604	OR4S1	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs1483118	11	48333226	OR4C3	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs1483121	11	48289936	OR4S1	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs1483123	11	48283349	OR4S1	flanking 5'UTR	0	0	-	0	5	0.063	0	5	0.027
rs1588069	11	48651447	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs4882001	11	48503251	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs4882088	11	48469518	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs4882129	11	48545350	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs603109	11	48400487	OR4A47	flanking 5'UTR	0	0	-	0	5	0.063	0	5	0.027
rs6485848	11	48498093	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs7129386	11	48573270	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs7395662	11	48475469	OR4A47	flanking 3'UTR	0	0	-	0	5	0.063	0	5	0.027
rs11136344	8	145131413	PARP10	coding	0	4	0.059	0	1	1	0	5	0.029
rs3936211	8	145133612	PARP10	flanking 5'UTR	0	4	0.059	0	1	1	0	5	0.029
rs6420181	8	145153236	SPATC1	flanking 5'UTR	0	4	0.059	0	1	1	0	5	0.029
rs1009780	6	124482442	TCBA1	intron	4	2	0.685	8	1	0.016	12	3	0.035
rs11758638	6	124496015	TCBA1	intron	4	2	0.685	8	1	0.016	12	3	0.035
rs10014421	4	57905392	IGFBP7	flanking 5'UTR	1	3	0.368	5	12	0.146	6	15	0.047
rs10488926	4	57907414	IGFBP7	flanking 5'UTR	1	3	0.368	5	12	0.146	6	15	0.047
rs12500068	4	57939482	IGFBP7	flanking 5'UTR	1	3	0.368	5	12	0.146	6	15	0.047
rs2079916	4	57913860	IGFBP7	flanking 5'UTR	1	3	0.368	5	12	0.146	6	15	0.047
rs2098159	4	57913664	IGFBP7	flanking 5'UTR	1	3	0.368	5	12	0.146	6	15	0.047

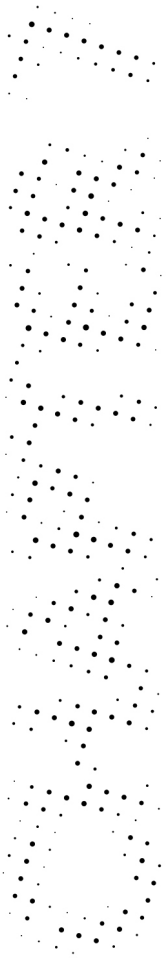
P values are calculated by the exact Cochran-Mantel-Haenszel test (for pooled populations) and the two-tailed Fisher exact test (for separate populations) and are not corrected for multiple testing

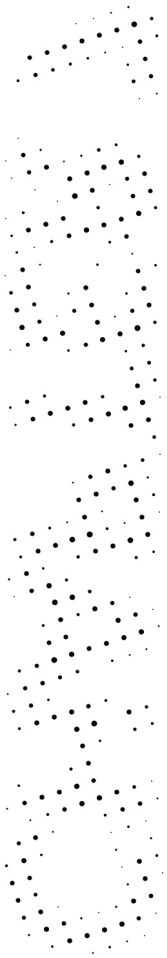
Supplementary Table 2 SNPs showing association with ALS by loss of copy number in the combined Irish and Dutch study populations, at uncorrected p values below 0.05

					IRELAND			THE NETHERLANDS			POOLED		
SNP	Chr.	Position	Gene	Position relative to gene	ALS patients (n)	Controls (n)	P	ALS patients (n)	Controls (n)	P	ALS patients (n)	Controls (n)	P
rs1603756	11	51171349	*OR4A5*	flanking 3'UTR	2	0	0.4988	15	1	0.0005	17	1	0.0001
rs4323853	11	51077585	*OR4A5*	flanking 3'UTR	1	0	1	15	1	0.0005	16	1	0.0003
rs8189141	11	51084722	*OR4A5*	flanking 3'UTR	1	0	1	15	1	0.0005	16	1	0.0003
rs10501331	11	51285867	*OR4A5*	flanking 5'UTR	0	0	-	12	0	0.0005	12	0	0.0005
rs527464	11	51297186	*OR4A5*	flanking 5'UTR	0	0	-	12	0	0.0005	12	0	0.0005
rs6816518	4	21738394	*GPR125*	flanking 3'UTR	1	11	0.0028	0	0	-	1	11	0.0028
rs4501212	4	21728213	*GPR125*	flanking 3'UTR	1	10	0.0053	0	0	-	1	10	0.0053
rs2554659	8	3776955	*CSMD1*	intron	9	24	0.0061	0	0	-	9	24	0.0061
rs2528691	7	110822399	*IMMP2L*	flanking 5'UTR	0	4	0.0592	0	3	0.1153	0	7	0.0068
rs10094534	8	15455979	*TUSC3*	intron	7	22	0.0035	10	13	0.5283	17	35	0.0098
rs1346590	8	15453141	*TUSC3*	intron	7	22	0.0035	10	13	0.5283	17	35	0.0098
rs7821056	8	15450330	*TUSC3*	intron	7	22	0.0035	10	13	0.5283	17	35	0.0098
rs882696	8	15447669	*TUSC3*	intron	7	22	0.0035	10	13	0.5283	17	35	0.0098
rs12671676	7	110706804	*IMMP2L*	flanking 3'UTR	0	4	0.0592	1	5	0.115	1	9	0.0101
rs2396393	7	110785151	*IMMP2L*	intron	0	4	0.0592	1	5	0.115	1	9	0.0101
rs4540344	7	110787188	*IMMP2L*	intron	0	4	0.0592	1	5	0.115	1	9	0.0101
rs2613596	7	110824195	*IMMP2L*	flanking 5'UTR	0	4	0.0592	0	2	0.2372	0	6	0.0140
rs11975331	7	110754307	*IMMP2L*	intron	0	5	0.029	2	5	0.2758	2	10	0.0198
rs2529489	7	110753847	*IMMP2L*	intron	0	5	0.029	2	5	0.2758	2	10	0.0198
rs2529490	7	110763178	*IMMP2L*	intron	0	5	0.029	2	5	0.2758	2	10	0.0198
rs2613585	7	110780644	*IMMP2L*	intron	0	5	0.029	2	5	0.2758	2	10	0.0198
rs2613588	7	110777561	*IMMP2L*	intron	0	5	0.029	2	5	0.2758	2	10	0.0198
rs4131198	8	43782832	*POTE8*	flanking 3'UTR	7	1	0.0677	6	2	0.2881	13	3	0.0207
rs8175516	8	43789367	*POTE8*	flanking 3'UTR	7	1	0.0677	6	2	0.2881	13	3	0.0207
rs1592593	11	50599126	*OR4C12*	flanking 5'UTR	1	0	1	10	2	0.038	11	2	0.0221
rs10882672	10	82875029	*SH2D4B*	flanking 3'UTR	1	2	0.6204	8	19	0.03	9	21	0.0248
rs1953661	10	82875955	*SH2D4B*	flanking 3'UTR	1	2	0.6204	8	19	0.03	9	21	0.0248
rs379800	10	82869699	*SH2D4B*	flanking 3'UTR	1	2	0.6204	8	19	0.03	9	21	0.0248
rs1110961	15	21920736	*NDN*	flanking 5'UTR	0	1	0.4951	0	4	0.056	0	5	0.0277
rs11629740	15	21927715	*NDN*	flanking 5'UTR	0	1	0.4951	0	4	0.056	0	5	0.0277
rs10822906	10	68167990	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs12109205	5	45912435	*HCN1*	flanking 5'UTR	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs12268982	10	68180377	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs12763151	10	68193332	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs13340341	5	45896693	*HCN1*	flanking 5'UTR	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs4311951	10	68190451	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs4331911	5	45895132	*HCN1*	flanking 5'UTR	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs4341430	10	68167094	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs4433471	10	68174741	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs4494211	10	68194973	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs7092091	10	68169242	*CTNNA3*	intron	0	2	0.2445	0	3	0.1153	0	5	0.0282
rs1543422	14	98252524	*FLJ25773*	intron	1	0	1	5	0	0.0622	6	0	0.0311
rs17097744	14	98262193	*FLJ25773*	flanking 3'UTR	1	0	1	5	0	0.0622	6	0	0.0311

rs1950489	14	98248060	*FLJ25773*	intron	1	0	1	5	0	0.0622	6	0	0.0311
rs2016624	14	98251139	*FLJ25773*	intron	1	0	1	5	0	0.0622	6	0	0.0311
rs12446298	16	77831605	*WWOX*	flanking 3'UTR	1	7	0.0357	0	0	-	1	7	0.0357
rs693764	16	77824930	*WWOX*	flanking 3'UTR	1	7	0.0357	0	0	-	1	7	0.0357
rs16959079	15	32530025	*GOLGA8A*	flanking 5'UTR	19	33	0.0374	0	0	-	19	33	0.0374
rs3812940	15	32575858	*GOLGA8B*	flanking 3'UTR	19	33	0.0374	0	0	-	19	33	0.0374
rs2872480	11	50578631	*OR4C12*	flanking 5'UTR	0	0	-	10	2	0.038	10	2	0.0380
rs10503137	18	64817882	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs1540432	18	64816131	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs17080075	18	64832896	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs2163474	18	64824572	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs3933569	18	64823011	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs4891718	18	64819792	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs8085022	18	64812093	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs930165	18	64827097	*C18orf14*	intron	3	0	0.2482	7	2	0.1787	10	2	0.0383
rs7423746	2	242653140	*FLJ40712*	flanking 3'UTR	16	21	0.3917	15	26	0.0565	31	47	0.0462

P values are calculated by the exact Cochran-Mantel-Haenszel test (for pooled populations) and the two-tailed Fisher exact test (for separate populations) and are not corrected for multiple testing.





A LARGE GENOME SCAN FOR RARE CNVS IN AMYOTROPHIC LATERAL SCLEROSIS

Human Molecular Genetics (2010)

Hylke M. Blauw,[1] Ammar Al-Chalabi,[3] 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], Fernando Rivadeneira,[7,8] Albert Hofman,[8] Andre G. Uitterlinden,[7,8] Lambertus A. Kiemeney,[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,[13,14,15,16] 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 Berg[1,25,*]

[1]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht 3584 CX, The Netherlands; [2]Department of Medical Genetics and Rudolf Magnus Institute, Utrecht 3584 CX, The Netherlands; [3]Department of Clinical Neuroscience, King's College London, Medical Research Council (MRC) Centre for Neurodegeneration Research, Institute of Psychiatry, London SE5 8AF, UK; [4]Department of Clinical Neuroscience, Umeå University, Umeå SE-901 85, Sweden; [5]Department of Neurology, University of Ulm, Ulm 89081, Germany; [6]Complex Genetics Section, Division of Biomedical Genetics, Department of Medical Genetics, University Medical Centre Utrecht, 3584 CX Utrecht, The Netherlands; [7]Department of Internal Medicine, Erasmus University Medical Centre Rotterdam, Rotterdam 3015 GE, The Netherlands; [8]Department of Epidemiology Erasmus University Medical Centre Rotterdam, Rotterdam 3015 GE, The Netherlands; [9]Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen 6500 HB, The Netherlands;

[10]Department of Neurology, Charité University Hospital, Humboldt-University, Berlin 13353, Germany; [11]Department of Neurology, Beaumont Hospital, Dublin, Ireland ; [12]Trinity College Institute of Neurosciences, Trinity College, Dublin, Ireland; [13]Department of Neurology, Massachusetts General Hospital, Charlestown, MA 02129, USA; [14]Department of Neurology, University of Massachusetts School of Medicine, Worcester, MA 01655, USA; [15]Howard Hughes Medical Institute, Cambridge, MA, 01605, USA; [16]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 01605, USA; [17]Department of Health Sciences and Department of Genetics, University of Leicester, Leicester LE1 7RH, UK; [18]Department of Neurology, Jagiellonian University, Krakow 31-007, Poland; [19]Experimental Neurology, University of Leuven, Leuven 3000, Belgium ; [20]Vesalius Research Centre, Flanders Institute for Biotechnology (VIB), Leuven 3000, Belgium ; [21]Department of Neurology, University Hospital Leuven, University of Leuven, Leuven 3000, Belgium; [22]Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Munich 80336, Germany; [23]Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and German Center for Neurodegenerative Diseases, Tübingen 72074, Germany; [24]UCLA Center for Neurobehavioral Genetics, Los Angeles 60095, USA; [25]These authors contributed equally to this work

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 *DPP6* locus (OR=3.59, $p=6.6x10^{-3}$), which has already been implicated in ALS pathogenesis, and the 15q11.2 locus, containing *NIPA1* (OR=12.46, $p=9.3x10^{-5}$), 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: *DPP6* locus: OR=1.92, p=0.097, pooled results: OR=2.64, $p=1.4x10^{-3}$; *NIPA1*: OR=3.23, p=0.041, pooled results: OR=6.20, $p=2.2x10^{-5}$). Our results highlight *DPP6* and *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 online 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. 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* 9.3×10^{-5}, Fisher exact test, OR = 12.46) (Tables 2 and S6). This region is also a known CNV region flanked by segmental duplications (t). 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* ($p= 9.4 \times 10^{-5}$, Fisher exact test, $p=2.6x10^{-3}$, Cochran-Mantel-Haenszel test) and slightly more modest for *DPP6* ($p= 9.1 \times 10^{-3}$, Fisher exact test, $p=1.2x10^{-2}$, Cochran-Mantel-Haenszel test) (not shown). Using the IBS-test as implented 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 assocation for the two loci: *NIPA1* ($p= 2.6 \times 10^{-3}$, 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 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.

Table 1. Summary of the study populations

	ALS cases	**Controls**	**Platform**
Genome-wide screen			
Belgium	272	276	Illumina CNV370
Ireland	204	205	Illumina 550K
Sweden	445	446	Illumina CNV370
The Netherlands	954	7,804	Illumina 300K, CNV370, 550K
Total	1,875	8,731	
Replication			
Germany	1,032	994	Illumina 300K, Fluidigm
Ireland	100	127	Illumina 610K
Scandinavian countries	441	-	Fluidigm
Poland	217	372	Fluidigm
The Netherlands	279	3,869	Illumina 550K, 610K, Fluidigm
United States	490	525	Illumina 300K
Total	2,559	5,887	
Combined	4,434	14,618	

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.

Locus	ALS cases		Controls		Fisher P [a]	OR [b]	CMH P [c]	OR [b]
	N	%	n	%				
Genome-wide screen:								
7q36.2:								
DPP6	10 of 1,875	0.53	13 of 8,731	0.15	$3.6x10^{-3}$	3.59	$6.6x10^{-3}$	3.70
15q11.2:								
TUBGCP5	7 of 1,875	0.37	4 of 8,731	0.05	$9.1x10^{-4}$	8.17	$1.8x10^{-2}$	5.29
CYFIP1	7 of 1,875	0.37	8 of 8,731	0.09	$9.1x10^{-3}$	4.09	$5.3x10^{-2}$	3.10
NIPA2	8 of 1,875	0.43	7 of 8,731	0.08	$1.9x10^{-3}$	5.34	$2.6x10^{-2}$	3.71
NIPA1	8 of 1,875	0.43	3 of 8,731	0.03	$9.3x10^{-5}$	12.46	$2.6x10^{-3}$	8.94
Replication:								
7q36.2:								
DPP6	10 of 2,559	0.39	12 of 5,887	0.20	$9.7 x10^{-2}$	1.92	0.24	1.71
15q11.2:								
TUBGCP5	6 of 2,559	0.23	7 of 5,887	0.12	0.17	1.97	0.22	2.21
CYFIP1	6 of 2,559	0.23	7 of 5,887	0.12	0.17	1.97	0.22	2.21
NIPA2	7 of 2,559	0.27	6 of 5,887	0.10	$4.2x10^{-2}$	3.05	0.13	2.77
NIPA1	7 of 2,559	0.27	5 of 5,887	0.08	$4.1x10^{-2}$	3.23	$5.5x10^{-2}$	4.15
Combined:								
7q36.2:								
DPP6	20 of 4,434	0.45	25 of 14,618	0.17	$1.4x10^{-3}$	2.64	$6.0x10^{-3}$	2.66
15q11.2:								
TUBGCP5	13 of 4,434	0.29	11 of 14,618	0.08	$1.0x10^{-3}$	3.90	$2.2x10^{-2}$	3.10
CYFIP1	13 of 4,434	0.29	15 of 14,618	0.10	$5.9x10^{-3}$	2.86	$4.1x10^{-2}$	2.48
NIPA2	15 of 4,434	0.34	13 of 14,618	0.09	$4.9 x10^{-4}$	3.81	$1.2x10^{-2}$	3.06
NIPA1	15 of 4,434	0.34	8 of 14,618	0.05	$2.2x10^{-5}$	6.20	$7.5x10^{-4}$	5.56

[a] One-tailed Fisher exact p value
[b] Odds ratio
[c] Cochran-Mantel-Haenszel p value (one-tailed), stratified by country of origin

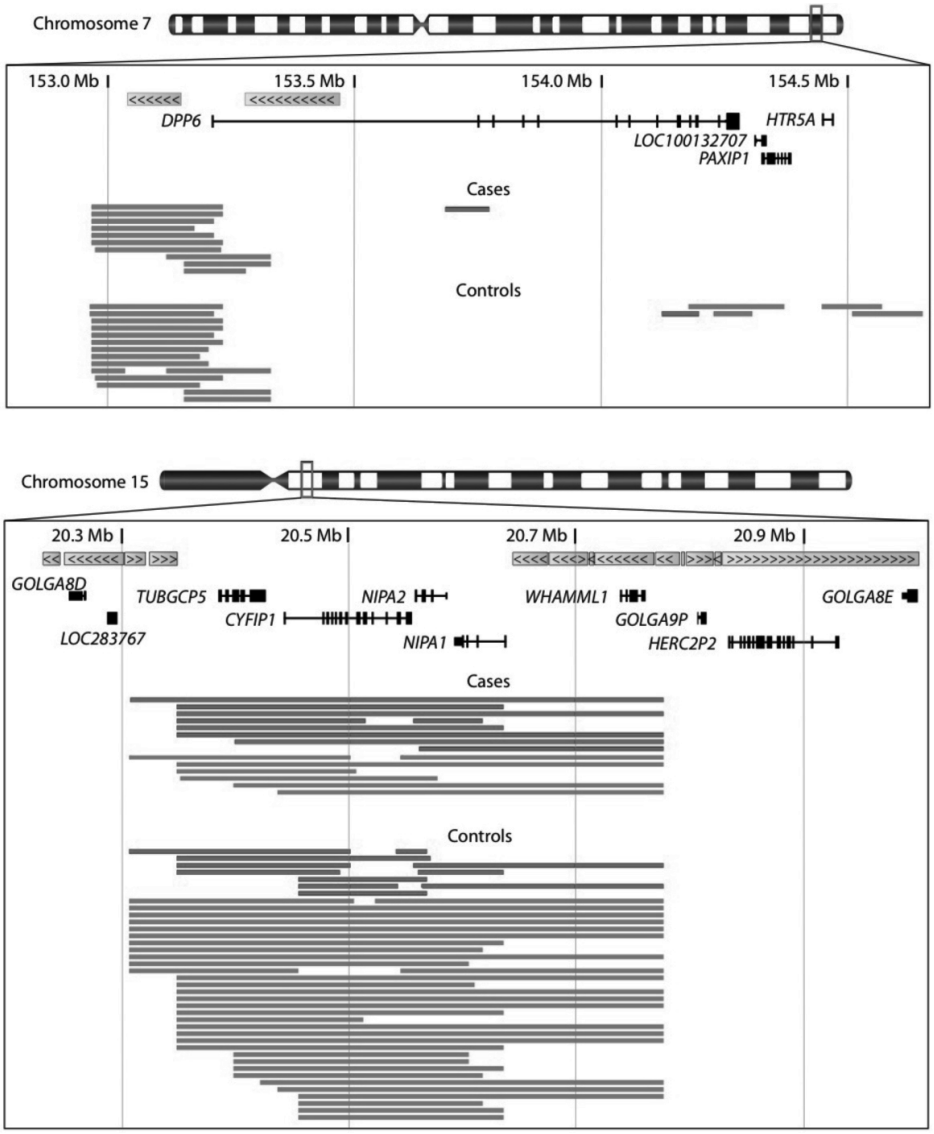


Figure 1. *Genomic organization of the DPP6 and 15q11.2 loci*
Figure shows duplications (light grey) and deletions (dark grey) observed in 1,875 ALS cases and 8,731 controls in the genome-wide screen. Segmental duplications are shown in light grey boxes with arrowheads.

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 assocations, because of differences in allele frequencies

of populations that are derived from differrent 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 aquire 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 ($r2 < 0.50$), 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 Prinses 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

1. Chiò, A., Logroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., Traynor, B.G. and Consortium, O.B.O.T.E. (2009) Prognostic factors in ALS: A critical review. *MALS*, 10, 310-323.
2. Dion, P.A., Daoud, H. and Rouleau, G.A. (2009) Genetics of motor neuron disorders: new insights into pathogenic mechanisms. *Nat. Rev. Genet.*, 10, 769-782.
3. Graham, A.J., Macdonald, A.M. and Hawkes, C.H. (1997) British motor neuron disease twin study. *J. Neurol. Neurosurg. Psychiatry*, 62, 562-569.
4. Van Es, M.A., Veldink, J.H., Saris, C.G.J., Blauw, H.M., Van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B. *et al.* (2009) Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.*, 41, 1083-1087.
5. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., Mccarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
6. Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y. *et al.* (2008) Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.*, 82, 477-488.
7. Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P.H., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E. *et al.* (2008) Large recurrent microdeletions associated with schizophrenia. *Nature*, 455, 232-236.
8. De Kovel, C.G.F., Trucks, H., Helbig, I., Mefford, H.C., Baker, C., Leu, C., Kluck, C., Muhle, H., Von Spiczak, S., Ostertag, P. *et al.* (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain*, 133, 23-32.
9. Blauw, H.M., Veldink, J.H., Van Es, M.A., Van Vught, P.W., Saris, C.G.J., van der Zwaag, B., Franke, L., Burbach, J.P.H., Wokke, J.H., Ophoff, R.A. *et al.* (2008) Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. *Lancet neurology*, 7, 319-326.

10. Cronin, S., Blauw, H.M., Veldink, J.H., Van Es, M.A., Ophoff, R.A., Bradley, D.G., Van Den Berg, L.H. and Hardiman, O. (2008) Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. *Hum. Mol. Genet.*, 30.
11. Wain, L.V., Pedroso, I., Landers, J.E., Breen, G., Shaw, C.E., Leigh, P.N., Brown, R.H., Tobin, M.D. and Al-Chalabi, A. (2009) The role of copy number variation in susceptibility to amyotrophic lateral sclerosis: genome-wide association study and comparison with published loci. *PLoS ONE*, 4, e8175.
12. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, 81, 559-575.
13. Meyer, T., Alber, B., Roemer, K., Martin, T., Kalscheuer, V.M., Göttert, E., Zang, K.D., Ludolph, A.C., Ropers, H.-H. and Prudlo, J. (2003) High rate of constitutional chromosomal rearrangements in apparently sporadic ALS. *Neurology*, 60, 1348-1350.
14. Locke, D.P., Sharp, A.J., Mccarroll, S.A., McGrath, S.D., Newman, T.L., Cheng, Z., Schwartz, S., Albertson, D.G., Pinkel, D., Altshuler, D.M. *et al.* (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. *Am. J. Hum. Genet.*, 79, 275-290.
15. Van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Van Den Bosch, L., De Jong, S.W., De Jong, V., Baas, F., Van't Slot, R. *et al.* (2008) Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. *Nat. Genet.*, 40, 29-31.
16. van Es, M.A., van Vught, P.W., van Kempen, G., Blauw, H.M., Veldink, J.H. and van den Berg, L.H. (2009) Dpp6 is associated with susceptibility to progressive spinal muscular atrophy. *Neurology*, 72, 1184-1185.
17. Cronin, S., Berger, S., Ding, J., Schymick, J.C., Washecka, N., Hernandez, D.G., Greenway, M.J., Bradley, D.G., Traynor, B.J. and Hardiman, O. (2008) A genome-wide association study of sporadic ALS in a homogenous Irish population. *Hum. Mol. Genet.*, 17, 768-774.
18. Del Bo, R., Ghezzi, S., Corti, S., Santoro, D., Prelle, A., Mancuso, M., Siciliano, G., Briani, C., Murri, L., Bresolin, N. *et al.* (2008) DPP6 gene variability confers increased risk of developing sporadic amyotrophic lateral sclerosis in Italian patients. *J. Neurol. Neurosurg. Psychiatry*, 79, 1085.
19. van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H. *et al.* (2009) Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.*, 41, 1083-1087.
20. Daoud, H., Valdmanis, P.N., Dion, P.A. and Rouleau, G.A. (2009) Analysis of DPP6 and FGGY as candidate genes for amyotrophic lateral sclerosis. *Amyotroph. Lateral. Scler.*
21. Fogh, I., D'Alfonso, S., Gellera, C., Ratti, A., Cereda, C., Penco, S., Corrado, L., Soraru, G., Castellotti, B., Tiloca, C. *et al.* (2009) No association of DPP6 with amyotrophic lateral sclerosis in an Italian population. *Neurobiol. Aging.*
22. Chio, A., Schymick, J.C., Restagno, G., Scholz, S.W., Lombardo, F., Lai, S.L., Mora, G., Fung, H.C., Britton, A., Arepalli, S. *et al.* (2009) A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.*, 18, 1524-1532.
23. Wong, W., Newell, E.W., Jugloff, D.G., Jones, O.T. and Schlichter, L.C. (2002) Cell surface targeting and clustering interactions between heterologously expressed PSD-95 and the Shal voltage-gated potassium channel, Kv4.2. *J Biol Chem*, 277, 20423-20430.
24. Feuk, L., Carson, A.R. and Scherer, S.W. (2006) Structural variation in the human genome. *Nat. Rev. Genet.*, 7, 85-97.
25. Rainier, S., Chai, J.-H., Tokarz, D., Nicholls, R.D. and Fink, J.K. (2003) NIPA1 gene mutations cause autosomal dominant hereditary spastic paraplegia (SPG6). *Am. J. Hum. Genet.*, 73, 967-971.
26. Zhao, J., Matthies, D.S., Botzolakis, E.J., Macdonald, R.L., Blakely, R.D. and Hedera, P. (2008) Hereditary Spastic Paraplegia-Associated Mutations in the NIPA1 Gene and Its Caenorhabditis elegans Homolog Trigger Neural Degeneration In Vitro and In Vivo through a Gain-of-Function Mechanism. *J. Neurosci.*, 28, 13938-13951.
27. Orlacchio, A., Babalini, C., Borreca, A., Patrono, C., Massa, R., Basaran, S., Munhoz, R.P., Rogaeva, E.A., St George-Hyslop, P.H., Bernardi, G. *et al.* (2010) SPATACSIN mutations cause autosomal recessive juvenile amyotrophic lateral sclerosis. *Brain*, 1-8.
28. Brugman, F., Wokke, J.H.J., Scheffer, H., Versteeg, M.H.A., Sistermans, E.A. and Van Den Berg, L.H. (2005)

Spastin mutations in sporadic adult-onset upper motor neuron syndromes. *Ann. Neurol.*, 58, 865-869.
29. Meyer, T., Schwan, A., Dullinger, J.S., Brocke, J., Hoffmann, K.-T., Nolte, C.H., Hopt, A., Kopp, U., Andersen, P., Epplen, J.T. *et al.* (2005) Early-onset ALS with long-term survival associated with spastin gene mutation. *Neurology*, 65, 141-143.
30. Tsang, H.T.H., Edwards, T.L., Wang, X., Connell, J.W., Davies, R.J., Durrington, H.J., O'Kane, C.J., Luzio, J.P. and Reid, E. (2009) The hereditary spastic paraplegia proteins NIPA1, spastin and spartin are inhibitors of mammalian BMP signalling. *Hum. Mol. Genet.*, 18, 3805-3821.
31. Williams, A.H., Valdez, G., Moresi, V., Qi, X., Mcanally, J., Elliott, J.L., Bassel-Duby, R., Sanes, J.R. and Olson, E.N. (2009) MicroRNA-206 Delays ALS Progression and Promotes Regeneration of Neuromuscular Synapses in Mice. *Science*, 326, 1549-1554.
32. Zollner, S. and Pritchard, J.K. (2007) Overcoming the winner's curse: estimating penetrance parameters from case-control data. *Am. J. Hum. Genet.*, 80, 605-615.
33. Van Es, M.A., Van Vught, P.W., Blauw, H.M., Franke, L., Saris, C.G., Andersen, P.M., Van Den Bosch, L., De Jong, S.W., van 't Slot, R., Birve, A. *et al.* (2007) ITPR2 as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet neurology*, 6, 869-877.
34. Schymick, J., Scholz, S., Fung, H., Britton, A., Arepalli, S., Gibbs, J., Lombardo, F., Matarin, M., Kasperaviciute, D. and Hernandez, D. (2007) Genome-wide genotyping in amyotrophic lateral sclerosis and neurologically normal controls: first stage analysis and public release of data. *The Lancet Neurology*, 6, 322-328.
35. Chiò, A., Schymick, J.C., Restagno, G., Scholz, S.W., Lombardo, F., Lai, S.-L., Mora, G., Fung, H.-C., Britton, A., Arepalli, S. *et al.* (2009) A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.*, 18, 1524-1532.
36. Dunckley, T., Huentelman, M.J., Craig, D.W., Pearson, J.V., Szelinger, S., Joshipura, K., Halperin, R.F., Stamper, C., Jensen, K.R., Letizia, D. *et al.* (2007) Whole-genome analysis of sporadic amyotrophic lateral sclerosis. *N. Engl. J. Med.*, 357, 775-788.
37. Brooks, B.R. (1994) El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. *J. Neurol. Sci.*, 124 Suppl, 96-107.
38. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R. and Witteman, J.C. (2007) The Rotterdam Study: objectives and design update. *Eur. J. Epidemiol.*, 22, 819-829.
39. Kiemeney, L.A., Thorlacius, S., Sulem, P., Geller, F., Aben, K.K.H., Stacey, S.N., Gudmundsson, J., Jakobsdottir, M., Bergthorsson, J.T., Sigurdsson, A. *et al.* (2008) Sequence variant on 8q24 confers susceptibility to urinary bladder cancer. *Nat. Genet.*, 40, 1307-1312.
40. Stefansson, H., Ophoff, R.A., Steinberg, S., Andreassen, O.A., Cichon, S., Rujescu, D., Werge, T., Pietiläinen, O.P.H., Mors, O., Mortensen, P.B. *et al.* (2009) Common variants conferring risk of schizophrenia. *Nature*, 1-5.
41. Simón-Sánchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G. *et al.* (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat. Genet.*, 41, 1308-1312.
42. Landers, J.E., Melki, J., Meininger, V., Glass, J.D., van den Berg, L.H., Van Es, M.A., Sapp, P.C., Van Vught, P.W.J., McKenna-Yasek, D.M., Blauw, H.M. *et al.* (2009) Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci. USA*, 106, 9004-9009.
43. Marioni, J.C., Thorne, N.P., Valsesia, A., Fitzgerald, T., Redon, R., Fiegler, H., Andrews, T.D., Stranger, B.E., Lynch, A.G., Dermitzakis, E.T. *et al.* (2007) Breaking the waves: improved detection of copy number variation from microarray-based comparative genomic hybridization. *Genome Biol.*, 8, R228.
44. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H. and Bucan, M. (2007) PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, 17, 1665-1674.
45. Itsara, A., Cooper, G., Baker, C., Girirajan, S., Li, J., Absher, D., Krauss, R., Myers, R., Ridker, P. and Chasman, D. (2009) Population Analysis of Large Copy Number Variants and Hotspots of Human Genetic Disease. *Am. J.*

Hum. Genet., 14.
46. Qin, J., Jones, R.C. and Ramakrishnan, R. (2008) Studying copy number variations using a nanofluidic platform. *Nucleic Acids Res.*, 36, e116-e116.
47. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. and Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.*, 8, R19.
48. Gonzalez, E. (2005) The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility. *Science*, 307, 1434-1440.
49. Scrucca, L. (2009) qcc R package. 1-7.
50. Mccarroll, S.A. and Altshuler, D.M. (2007) Copy-number variation and association studies of human disease. *Nat. Genet.*, 39, S37-S42.
51. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21, 263-265.
52. Faul, F., Erdfelder, E., Lang, A.G. and Buchner, A. (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods*, 39, 175-191.

SUPPLEMENTARY MATERIAL

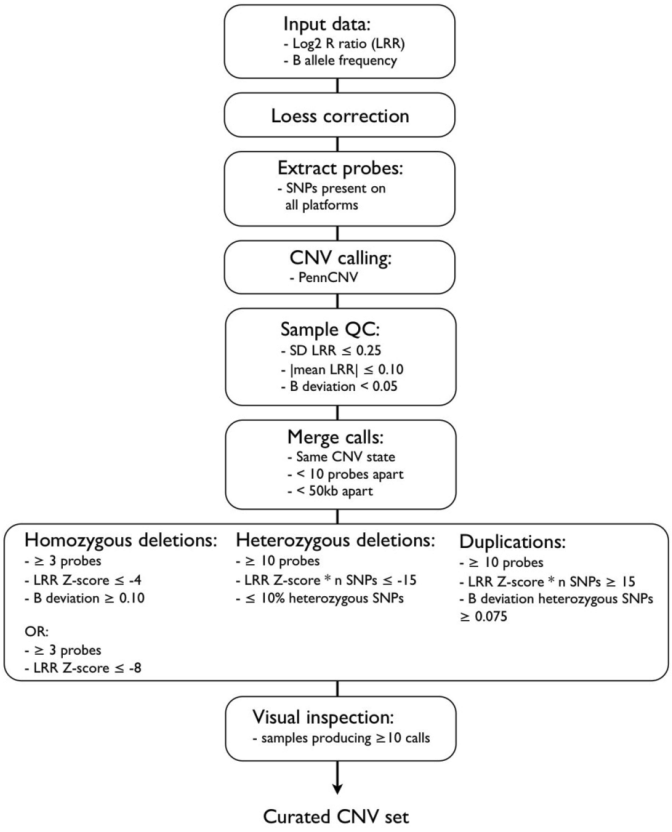


Figure S1. CNV calling procedure

Log2 R Ratio and B allele frequency metrics were used as input for CNV calling procedures. Because different platforms were used, only probes were used that were present on all arrays. The applied quality control filters have been described by Itsara et al.(45). While Itsara et al. used a threshold Z-score of -1.5 and 1.5 for heterozygous deletions and duplications, we used the product of the Z-score and the number of affected probes, creating more stringency for small CNVs and less stringency for large CNVs.

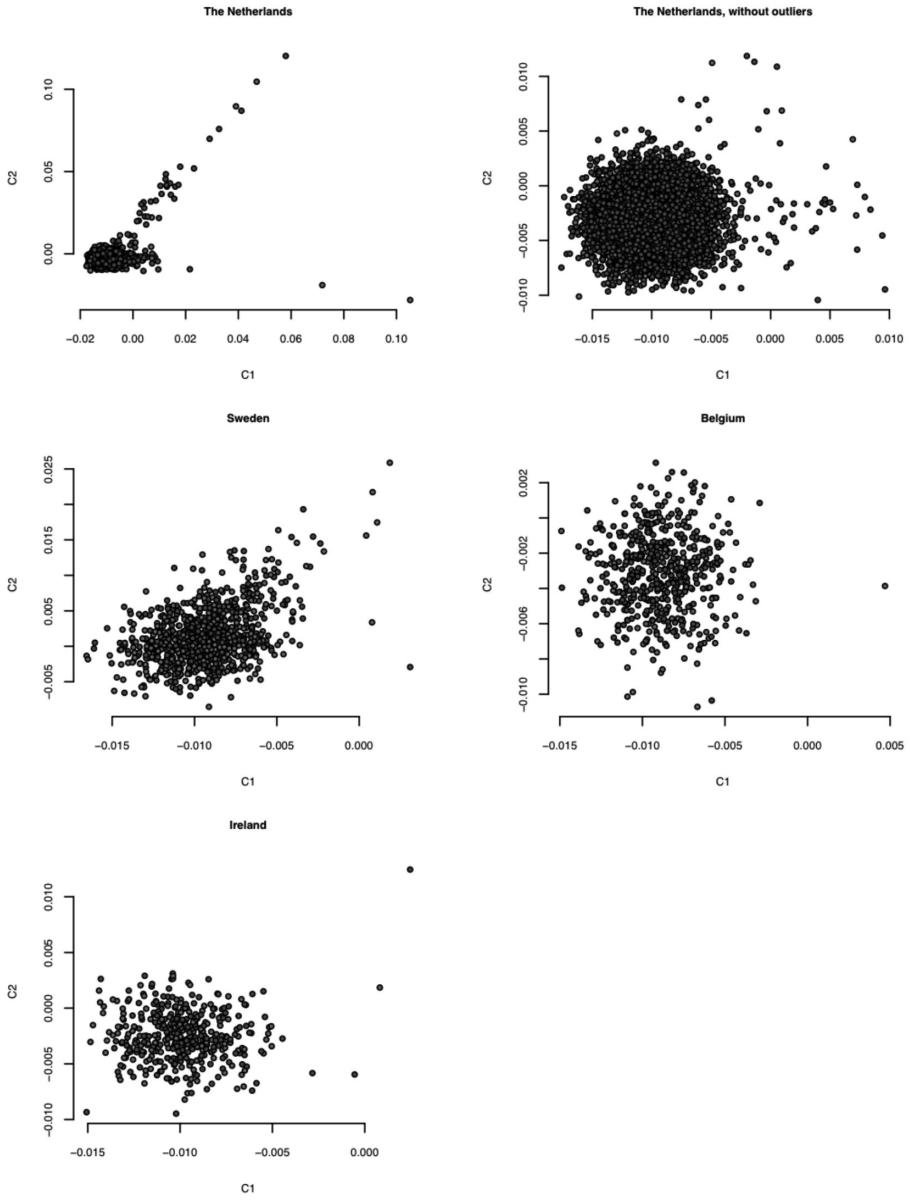


Figure S2. MDS plots per country.
The first two components of the multi-dimensional scaling analysis are plotted for each country in the discovery phase. For the Dutch, a plot is shown with and without 35 outliers (see main text for details). ALS cases are shown in red, controls are shown in blue.

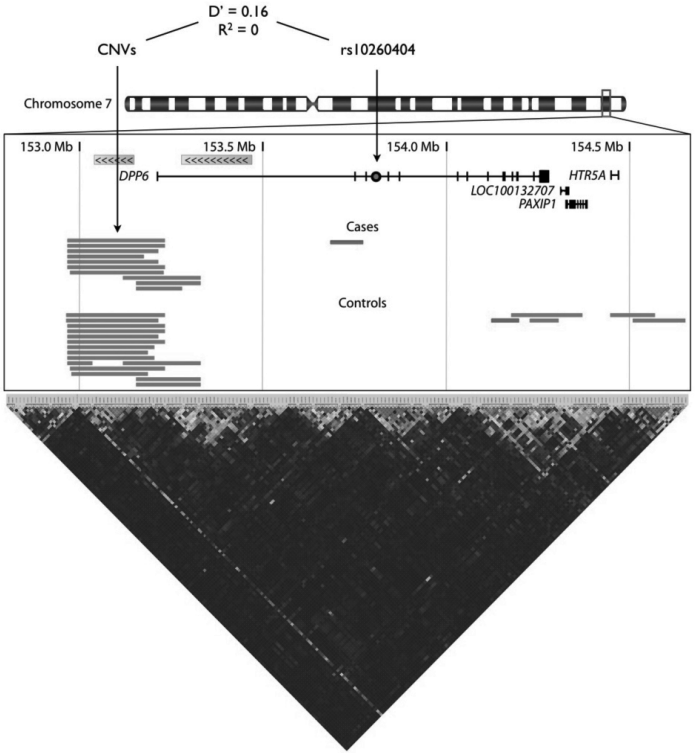


Figure S3. *LD structure of the DPP6 locus*
Figure shows the linkage disequilibrium structure of the DPP6 locus, the CNVs found in the genome-wide screen and the position of SNP rs10260404, which was found to be associated with ALS (15, 17, 18). LD structure was calculated on the individuals in the genome-wide screen.

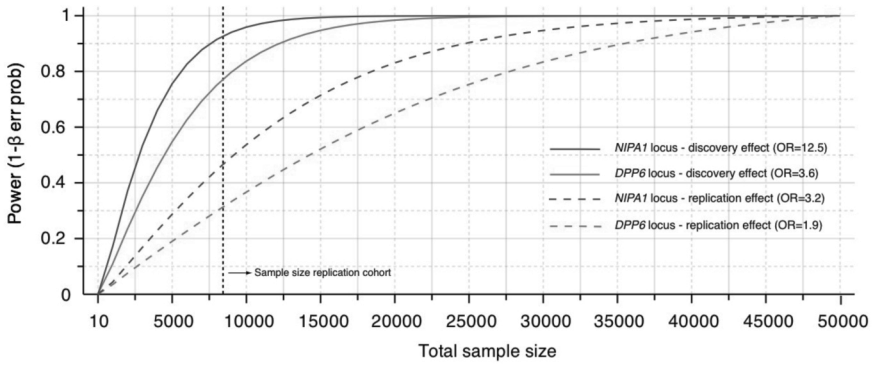


Figure S4. *Power calculations for the replication cohort*
Power to detect a true association in the replication cohort was calculated at α=0.025 for the effect sizes found in the genome-wide screen (continuous lines) and for the effect sizes found in the replication study (dashed lines). The vertical dashed line depicts the sample size of the replication cohort.

Table S1. Breakdown of included datsets after applied quality control.

	Source	ALS cases	Controls	Platform
Genome-wide screen:				
The Netherlands	University Hospital, Utrecht [a]	415	403	Illumina 300K
The Netherlands	University Hospital, Utrecht [b]	539	578	Illumina CNV370
The Netherlands	Rotterdam Study, RS1 (ERGO) [c]	-	4,771	Illumina 550K
The Netherlands	Radboud University, Nijmegen [d]	-	1,622	Illumina CNV370
The Netherlands	University Hospital, Utrecht [e]	-	430	Illumina 550K
Belgium	University Hospital, Leuven [b]	272	276	Illumina CNV370
Ireland	Beaumont Hospital, Dublin [f]	204	205	Illumina 550K
Sweden	University Hospital, Umeå [b]	445	446	Illumina CNV370
Total		1,875	8,731	
Replication:				
The Netherlands	University Hospital, Utrecht	279	908	Fluidigm
The Netherlands	Rotterdam Study, RS2 [c]	-	944	Illumina 550K
The Netherlands	Rotterdam Study, RS3 [c]	-	2,017	Illumina 610K
United States	Massachusetts General Hospital, Boston; Emory University Hospital, Atlanta [g]	490	525	Illumina 300K
Germany	University Hospital, Ulm; Charité University Hospital, Berlin	1,032	684	Fluidigm
Germany	University of Munich [e]	-	310	Illumina 300K
Ireland	Beaumont Hospital, Dublin	100	127	Illumina 610K
Poland	Jagiellonian University, Krakow	217	372	Fluidigm
Scandinavian countries	University Hospital, Umeå	441	-	Fluidigm
Total		2,559	5,887	
Combined		4,434	14,618	

[a] GWAS data and CNV data have been published previously (9, 15).
[b] Genotyped as part of an ALS GWAS. More information available in ref. (4)
[c] More detailed information is available in ref. (38)
[d] More detailed information is available in ref. (39)
[e] GWAS and CNV data have been published previously (7, 40)
[f] Genotyped as part of an Irish ALS GWAS.(17) CNV data have been published previously (10).
[g] GWAS and CNV data have been published previously (11, 42).

Table S2. CNV characteristics.

All CNVs	Duplications and deletions			Duplications			Deletions		
	SALS cases	Controls	*P*	SALS cases	Controls	*P*	SALS cases	Controls	*P*
N CNVs	2,951	13,552		1,580	7,054		1,371	6,498	
Average CNVs per individual	1.57	1.55	0.28	0.84	0.81	0.08	0.73	0.74	0.69
Proportion of samples with ≥1 CNV	0.76	0.75	0.18	0.55	0.53	0.04	0.48	0.48	0.55
Average CNV length (kilobase)	154.9	158.4	0.84	176.7	181	0.79	135.6	132.8	0.23
Genes spanned by CNVs	2.08	2.30	0.99	1.407	1.527	0.94	0.68	0.78	0.96
CNVs spanning ≥1 gene	0.57	0.57	0.45	0.43	0.41	0.08	0.25	0.27	0.90
Genes per kilobase	0.01	0.01	0.99	0.01	0.01	1.00	$9.1x10^{-3}$	0.01	0.84
Non-DGV genes spanned by CNVs	0.38	0.42	0.91	0.21	0.24	0.81	0.16	0.19	0.86
CNVs spanning ≥1 non-DGV gene	0.17	0.17	0.63	0.10	0.10	0.18	0.07	0.08	0.90
Non-DGV genes per kilobase	$1.6x10^{-3}$	$1.8x10^{-3}$	0.85	$1.7x10^{-3}$	$1.6x10^{-3}$	0.35	$2.1x10^{-3}$	$1.9x10^{-3}$	0.27
Large CNVs (>500kb)	**Duplications and deletions**			**Duplications**			**Deletions**		
	SALS cases	Controls	*P*	SALS cases	Controls	*P*	SALS cases	Controls	*P*
N CNVs	78	476		51	343		27	133	
Average CNVs per individual	0.04	0.05	0.99	0.03	0.04	0.99	0.01	0.02	0.63
Proportion of samples with ≥1 CNV	0.03	0.05	0.99	0.03	0.04	0.99	0.01	0.01	0.63
Average CNV length (kilobase)	665.1	720.5	0.98	660.7	722.7	0.95	676.3	714.5	0.87
Genes spanned by CNVs	0.14	0.18	0.85	0.12	0.15	0.79	0.02	0.03	0.73
CNVs spanning ≥1 gene	0.03	0.03	0.92	0.02	0.03	0.96	$8.0x10^{-3}$	$6.6x10^{-3}$	0.30
Genes per kilobase	$4.8x10^{-3}$	$4.8x10^{-3}$	0.51	$5.9x10^{-3}$	$5.4x10^{-3}$	0.30	$2.3x10^{-3}$	$3.3x10^{-3}$	0.67
Non-DGV genes spanned by CNVs	0.04	0.05	0.79	0.03	0.04	0.75	$6.4x10^{-3}$	$9.4x10^{-3}$	0.67
CNVs spanning ≥1 non-DGV gene	0.01	0.02	0.79	0.01	0.01	0.80	$3.2x10^{-3}$	$3.5x10^{-3}$	0.66
Non-DGV genes per kilobase	$1.5x10^{-3}$	$1.3x10^{-3}$	0.29	$1.8x10^{-3}$	$1.5x10^{-3}$	0.19	$7.3x10^{-4}$	$1.0x10^{-3}$	0.60

Characteristics of CNVs in the genome-wide screen in 1,875 SALS cases and 8,731 controls. P values are calculated from 10,000 permutations. "Non-DGV genes" are genes not spanned by CNVs in the Database of Genomic Variants (excluding BAC-array studies).

Table S4. Results for the genome-wide screen.

		All CNVs			Duplications only			Deletions only		
Locus	**RefSeq ID**	**ALS cases N=1,875**	**Controls N=8,731**	***P***	**ALS cases N=1,875**	**Controls N=8,731**	***P***	**ALS cases N=1,875**	**Controls N=8,731**	***P***
7q36.2	*DPP6*	10 (0.53%)	13 (0.15%)	$3.6x10^{-3}$	9 (0.48%)	12 (0.14%)	$6.3x10^{-3}$	1 (0.05%)	1 (0.01%)	0.32
15q11.2	*TUBGCP5*	12 (0.64%)	31 (0.36%)	$6.5 x10^{-2}$	5 (0.27%)	27 (0.31%)	0.69	7 (0.37%)	4 (0.05%)	$9.1x10^{-4}$
	CYFIP1	13 (0.69%)	42 (0.48%)	0.16	6 (0.32%)	34 (0.39%)	0.73	7 (0.37%)	8 (0.09%)	$9.1 x10^{-3}$
	NIPA2	13 (0.69%)	38 (0.44%)	0.10	5 (0.27%)	31 (0.36%)	0.79	8 (0.43%)	7 (0.08%)	$1.9 x10^{-3}$
	NIPA1	12 (0.64%)	34 (0.39%)	0.10	4 (0.21%)	31 (0.36%)	0.89	8 (0.43%)	3 (0.03%)	$9.3x10^{-5}$
19q13.2	*PLEKHG2*	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$	-	-	-	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$
	RPS16	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$	-	-	-	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$
	SUPT5H	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$	-	-	-	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$
	TIMM50	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$	-	-	-	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$
	DLL3	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$	-	-	-	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$
	SELV	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$	-	-	-	3 (0.16%)	0 (0%)	$5.5 x10^{-3}$

Loci containing genes affected by CNVs with a frequency of <1% in controls and $p < 0.01$ after testing with all CNVs, duplications only or deletions only. P values are from one-tailed Fisher exact tests. Counts per gene are shown with frequency in parentheses. The 3 CNV calls in the 19q13.2 locus did not validate. All tested CNVs in the remaining loci (25 of 25) were experimentally validated.

Table S5. *DPP6* CNVs.

	Cases	Controls	Cases		Controls		Fisher P [a]	OR	CMH P [b]	OR
			n	%	n	%				
Discovery phase										
The Netherlands	954	7,804	6	0.63	11	0.14				
Sweden	445	446	3	0.67	2	0.45				
Belgium	272	276	1	0.37	0	0.00				
Ireland	204	205	0	0.00	0	0.00				
Total Discovery phase	1,875	8,731	10	0.53	13	0.15	3.6×10^{-3}	3.59	6.6×10^{-3}	3.70
Replication phase										
The Netherlands	279	3,869	2	0.72	7	0.18				
United States	490	525	2	0.41	2	0.38				
Germany	1,032	994	3	0.29	1	0.10				
Ireland	100	127	0	0.00	0	0.00				
Poland	217	372	0	0.00	2	0.54				
Scandinavian countries	441	-	3	0.68	-	-				
Total Replication phase	2,559	5,887	10	0.39	12	0.20	9.7×10^{-2}	1.92	0.24	1.71
Combined	4,434	14,618	20	0.45	25	0.17	1.4×10^{-3}	2.64	6.0×10^{-3}	2.66

Results for the DPP6 locus, per country. Results are for all CNVs found at this locus (i.e. duplications and deletions combined).
[a] One-tailed Fisher exact test
[b] One-tailed Cochran-Mantel-Haenszel test, stratified by country of origin

Table S6. *NIPA1* deletions.

	Cases	Controls	Cases		Controls		Fisher *P* [a]	OR	CMH *P* [b]	OR
			n	%	n	%				
Discovery phase										
The Netherlands	954	7,804	2	0.21	3	0.04				
Sweden	445	446	3	0.67	0	0.00				
Belgium	272	276	2	0.74	0	0.00				
Ireland	204	205	1	0.49	0	0.00				
Total Discovery phase	1,875	8,731	8	0.43	3	0.03	9.3×10^{-5}	12.46	2.6×10^{-3}	8.94
Replication phase										
The Netherlands	279	3,869	2	0.72	3	0.08				
United States	490	525	1	0.20	0	0.00				
Germany	1,032	994	3	0.29	2	0.20				
Ireland	100	127	0	0.00	0	0.00				
Poland	217	372	0	0.00	0	0.00				
Scandinavian countries	441	-	1	0.23	-	-				
Total Replication phase	2,559	5,887	7	0.27	5	0.08	4.1×10^{-2}	3.23	5.5×10^{-2}	4.15
Combined	4,434	14,618	15	0.34	8	0.05	2.2×10^{-5}	6.20	7.5×10^{-4}	5.56

Results for the *NIPA1* locus, per country. Results are for deletions found in this locus.
[a] One-tailed Fisher exact test
[b] One-tailed Cochran-Mantel-Haenszel test, stratified by country of origin

Table S7. *DPP6* CNVs (top) and *NIPA1* deletions (bottom) found in the discovery phase, per country, per platform.

	300K	370K	550K
The Netherlands			
Cases	3	3	0
Controls	2	2	7
Sweden			
Cases	0	3	0
Controls	0	2	0

Belgium			
Cases	0	1	0
Controls	0	0	0
Ireland			
Cases	0	0	0
Controls	0	0	0

	300K	370K	550K
The Netherlands			
Cases	0	2	0
Controls	0	1	2
Sweden			
Cases	0	3	0
Controls	0	0	0
Belgium			
Cases	0	2	0
Controls	0	0	0
Ireland			
Cases	0	0	1
Controls	0	0	0

Table S8. Results for IBS tests for each population in the discovery phase.

Country	Cases	Controls	P IBS-test T1	P IBS-test T2
The Netherlands	954	7804	0.11	0.89
Belgium	272	276	0.09	0.91
Ireland	204	205	0.85	0.15
Sweden	445	446	$1.0x10^{-5}$	1

Table S9. Clinical characteristics of SALS cases with 15q11.2 deletions.

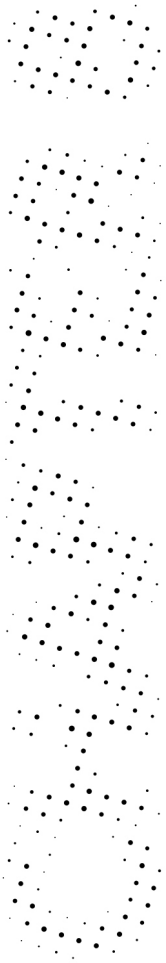
Patient	Country	Gender	Age onset	Site onset	Phenotype	Survival (Months)	Psychiatric co-morbidity	Migraine/ epilepsy
Patient A	Belgium	Male	53	Spinal	Classic ALS	43	No	Unknown
Patient B	Belgium	Male	59	Spinal	Classic ALS	14	No	Unknown
Patient C	Germany	Male	33	Spinal	Classic ALS	252	Unknown	Unknown
Patient D	Germany	Male	70	Spinal	Classic ALS	36	No	No
Patient E	Germany	Female	73	Spinal	LMN onset	14	Unknown	Unknown
Patient F	Ireland	Male	78	Spinal	Classic ALS	6	Possible FTD	No
Patient G	Norway	Male	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
Patient H	Sweden	Male	66	Spinal	LMN predominant	10	No	Short period of epileptic episodes
Patient I	Sweden	Male	58	Spinal	UMN predominant	39	No	No
Patient J	Sweden	Male	50	Spinal	Classic ALS	46	No	No
Patient K	The Netherlands	Male	70	Spinal	Classic ALS	20	No	No
Patient L	The Netherlands	Male	48	Spinal	Classic ALS	16	Unknown	Unknown
Patient M	The Netherlands	Female	67	Spinal	LMN onset	19	No	No
Patient N	The Netherlands	Female	70	Spinal	Classic ALS	29	No	No
Patient O	USA	Female	79	Spinal	Classic ALS	32	No	No

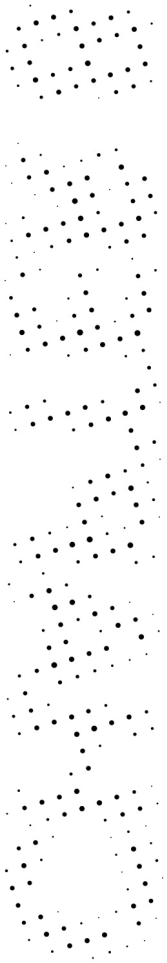
LMN: Lower motor neuron
UMN: Upper motor neuron
FTD: Fronto-temporal dementia
n.a.: Not available

Table S10. QPCR assay details.

Locus	Assay ID	Type	Chromosome	Position
7q36.2	Hs03643742_cn	Target assay	7	153,031,806
7q36.2	Hs04332196_cn	Target assay	7	153,060,849
7q36.2	Hs03643863_cn	Target assay	7	153,168,040
7q36.2	Hs04330326_cn	Target assay	7	153,276,294
7q36.2	Hs03648346_cn	Target assay	7	153,285,293
7q36.2	Hs01889186_cn	Target assay	7	153,774,290
7q36.2	Hs04976632_cn	Target assay	7	154,225,119
7q36.2	Hs00223125_cn	Target assay	7	154,276,435
15q11.2	Hs02859687_cn	Target assay	15	20,387,566
15q11.2	Hs02705800_cn	Target assay	15	20,507,736
15q11.2	Hs02979691_cn	Target assay	15	20,557,735
15q11.2	Hs00057491_cn	Target assay	15	20,600,411
19q13.2	Hs00353892_cn	Reference assay	19	44,672,396
14q11.2	RNase P Control Reagent	Reference assay	14	19,881,070

Genomic positions are derived from NCBI genome build 36.





COPY-NUMBER POLYMORPHISMS IN AMYOTROPHIC LATERAL SCLEROSIS

In preparation

Hylke M. Blauw,[1] Wouter van Rheenen,[1] Christopher Barnes,[2] Paul W. J. van Vught,[1] Lambertus A. Kiemeney,[3] Sita H.M. Vermeulen,[3] Claudia Schulte,[4] Thomas Gasser,[4] , Barbara Tomik,[5] Agnieszka Slowik,[5] Wim Robberecht,[6] Albert C. Ludolph,[7] Peter M. Andersen,[7,8] Jan H. Veldink, MD,[1#] Leonard H. van den Berg,[1#,*]

[1]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [2] Centre for Bioinformatics, Division of Molecular Biosciences and Institute of Mathematical Sciences, Imperial College London, London, UK; [3]Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, Nijmegen 6500 HB, The Netherlands; [4]Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and German Center for Neurodegenerative Diseases, Tübingen, Germany; [5]Department of Neurology, Jagiellonian University, Krakow, Poland; [6]Department of Neurology, University Hospital Leuven, University of Leuven, Leuven, Belgium and Vesalius Research Center, VIB, Leuven, Belgium; [7]Department of Neurology, University of Ulm, Ulm, Germany; [8]Department of Clinical Neuroscience, Umeå University, Sweden

[#] joint last authors

ABSTRACT

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease with unknown aetiology. The contribution of genetic risk factors is well recognised. Previous studies have aimed to identify genetic risk loci using genome-wide studies targeting SNPs and rare copy number variants (CNVs). While many examples of pathogenic CNVs exist in complex traits, including ALS, the role of polymorphic CNVs (copy number polymorphisms, CNPs) in ALS has not been systematically studied. To identify CNPs that predispose to ALS or influence its disease course, we did a CNP association study including a total of 3,441 ALS patients and 5,249 controls. In the first stage of the study we used the probe hybridisation intensities from Illumina CNV370 SNP arrays to type a set of previously described CNPs. From 253 successfully typed CNPs, we selected five that nominally ($p \leq 0.01$) associated to ALS susceptibly (n=4) or disease duration (n=1) and tested these in an independent cohort using qPCR and fragment analysis. In the replication cohort no CNP showed significant association. Notably, only one of the five CNPs was tagged by a SNP that was included in previous GWAS, which illustrates the additional value of this study. CNP association studies are technically challenging and in our current study we could type only a subset of known CNPs across the human genome. Given the observation that many CNPs have not been captured by previous GWASs and the current study, the systematic characterisation of CNPs in ALS may yield new insights into the genetic underpinnings of ALS.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that primarily affects motor neurons in the spinal cord and the motor cortex. The demise of motor neurons leads to progressive muscle weakness and spasticity, ultimately leading to respiratory insufficiency and death. The incidence is about 2-3 per 100,000 and the lifetime risk of developing the disease is 1:350-400. (1) The median survival is approximately 30 months after the onset of symptoms. (1) The only proven therapy to slow disease progression is with riluzole, a glutamate inhibitor, which improves survival by about 3 months. (2)

About 5-10% of ALS patients have a family history of the disease, usually with an autosomal dominant mode of inheritance. Several causative genes have been identified thus far, including *SOD1*, *TDP43*, *FUS*, *ANG* and *C9ORF72*. (3) In more than 50% of familial cases a genetic cause can now be identified. (4) However, the majority of ALS patients (90-95%) has no family history and are said to be sporadic. The mechanisms that contribute to sporadic ALS are largely unknown. The pathogenesis is considered to be multifactorial, with genetic and environmental risk factors contributing to disease susceptibility. The role of genetic risk factors is emphasised by the relatively high heritability, demonstrated in twin studies. (5) So far, only a few well-established loci have been identified that convey an increased risk for sporadic ALS, including *UNC13A*, *ATXN2* and *C9ORF72*. (6-9)

In addition to genome-wide association studies (GWAS), using single nucleotide polymorphisms (SNPs) as a means to identify associated loci, others and we have examined the role of copy-number variants (CNVs, insertions and deletions of genomic sequence) in ALS pathogenesis. (10-12) These studies aimed to identify deletions or duplications that associate with ALS susceptibility using the intensity data from the

GWAS SNP genotyping experiments. As a consequence of the used methodology that requires intensity deviations of several (typically >10) consecutive probes, these studies focused primarily on relatively large and (consequently) relatively rare genomic events. These methods are biased against smaller and more common CNVs: copy number polymorphisms (CNPs). These are CNVs that are polymorphic in the population: they are generally ancestral mutations and occur with a minor allele frequency of more than 1% in the general population. (13) Several complex traits were shown to be associated with CNPs including psoriasis, glomerulonephritis, Crohn's disease and body mass index. (14-18) Others and we have reported the association of ALS with a CNP containing the *SMN* (Survival of Motor Neuron) gene. (19) Apart from the *SMN* locus, the role of CNPs in ALS has not yet been studied systematically.

Although many CNPs are in linkage disequilibrium (LD) with neighbouring SNPs and are therefore indirectly captured in conventional SNP GWAS, a substantial part (40-50%) of CNPs is not captured by tagging SNPs on older genotyping arrays. (13) This is in part explained by the low probe coverage in CNP regions on the first generation genotyping platforms. More recently developed SNP arrays have included additional sets of markers to assay these gaps, enabling the direct interrogation of CNPs for association studies. Given the large proportion of "missed" CNPs by conventional SNP GWAS, the direct typing of CNPs is expected to be of additional value.

To examine the role of CNPs in ALS pathogenesis we did a genome-wide CNP association study using SNP intensity data, which have been previously generated for a SNP GWAS in ALS using the Illumina CNV370 array, which includes probes specifically targeted at CNP regions. Instead of previous studies (10-12) that employed a "per-sample" analysis method, we used a "per-CNP" method (using the probe intensity distribution of one probe across all samples), enabling the typing of smaller and more common CNPs. We did an initial genome-wide discovery phase and sought to replicate the most significantly associated CNPs in an independent population.

MATERIALS AND METHODS

Populations

ALS patients included in this study were diagnosed in specialised referral centers for neuromuscular diseases. Patients with a family history of ALS were excluded. The relevant medical ethical committees approved all studies, and all participants provided informed consent. The study populations are summarised in Table 1.

In the first (discovery) stage we used intensity data that was generated previously for a genome-wide association study in ALS using Illumina CNV370 BeadChips. These samples include ALS patients and controls from The Netherlands, Sweden and Belgium and have been described elsewhere in detail. (9) The included samples had passed quality control measures in the described study. In the final analyses we included 1,350 ALS patients and 3,262 controls.

In a second (replication) stage we used DNA from ALS patients and controls from The Netherlands, Germany and Poland. (9) The Dutch patients and controls have been recruited as part of a nationwide prospective study on motor neuron diseases in The Netherlands. The German ALS patients were recruited in the Department of Neurology, Ulm University, Ulm, and Charité University Hospital, Berlin. German controls were obtained from a prospective cohort study for Parkinson's disease in Tuebingen. (20) Polish ALS patients were recruited in the Department of Neurology of Jagiellonian University, Krakow. Polish controls were from Southern Poland, and were free of neurological disease.

Table 1. Study populations.

	ALS	Controls	Platform
Genome-wide			
Belgium	301	339	Illumina CNV370
Sweden	471	480	
The Netherlands	578	2443	
Total	1,350	3,262	
Replication			
Susceptibility			
Germany	1201	746	TaqMan qPCR (Fuidigm)
Poland	228	471	
The Netherlands	662	770	
Total	2,091	1,987	
Survival			
Germany	290	-	Fragment analysis
The Netherlands	677	-	
Total	1,067	-	

Genome-wide data analysis

Genome-wide intensity data were derived from hybridisations on Illumina CNV370 BeadChips. The genotyping experiments were carried out at Decode Genetics, Reykjavik, Iceland. R values (a normalized intensity metric) were extracted from Beadstudio software and quantile-normalised using Normtools (http://cnvtools.sourceforge.net/Normtools.html). Using the coordinates of the CNPs reported by McCarroll et al. (13) we then extracted markers that mapped to the genomic interval of each CNP. The quantile-normalized R values were used as input signal for further analyses. When CNPs were assayed by multiple markers, we used a summary measurement of multiple markers: either by principal component analysis or linear discriminant function, as described by Barnes et al. (21) We manually inspected the plots of the intensity distributions to determine the most likely number of components (copy number states) for each CNP. We then fitted Gaussian mixture models to the intensity distribution using CNVtools in R (version 2.10.1). In a reiterative process, a plot of each fit was inspected and if necessary refitted with different parameters until a correct fit was achieved. Association analyses were then carried out as described below.

Quantitative PCR

For replication and validation purposes, we did quantitative PCR (qPCR) experiments on the Fluidigm platform using BioMark 96.96 Dynamic Arrays (Fluidigm Corporation) using TaqMan Copy Number Assays (Applied Biosystems), similar to as described previously. (10) For each CNP to be tested, we selected two TaqMan Copy Number Assays. We used an assay targeting *TIMM50* as a reference. (10) Used TaqMan assays can be found in Supplementary Table 1. After 14 cycles of specific target amplification (STA), PCRs were done using a standard TaqMan protocol. Each PCR plate included a no-template

control and a standard curve of serially-diluted DNA. Each reaction was done in triplicate. Ct values were calculated with Fluidigm Real-Time PCR Analysis software with standard parameters. Because we conducted mutiple PCR assays in parallel, we could filter out low quality samples: we removed samples with a failure rate of > 80% of assays. We calculated estimated copy-numbers with qBase software (22), by normalisation against the reference gene, corrected for plate and assay specific amplification efficiencies. We then used these values as input signal for further analysis with CNVtools. (21)

Fragment analysis
One CNP (mc_10311) was tested in the replication experiment using fragment analysis with fluorescent-labeled primers. The SNP intensity data and publicly available data ((13), Database of Genomic Variants) showed that this CNP was compatible with a diallelic insertion/deletion polymorphism. We sequenced the breakpoints and developed a multiplex PCR to amplify both alleles in one reaction. We used one 6FAM-labeled forward primer (TCAAGCTGGTTGAAGAAATG), one reverse primer for amplification of the non-deletion allele (TCAATTCCAACTTCTTGCTATG) and one reverse primer to amplify the deletion allele (CCTCTAGACCACAGGCAATC). The resulting fragments are 239 and 296 bp and can be separated using fragment analysis based on their length. We used a standard touchdown protocol for PCR amplification (reaction conditions are available upon request). The resulting PCR product was then diluted and was loaded with HiDi formamide and a ROX GS500 ladder on an ABI 3730XL capillary sequencer (Applied biosystems). Fragment analysis was then done with GeneMapper software (version 3.7) (Applied Biosystems).

Statistical analysis
For association analyses we used multivariate logistic regression. To compute the most likely copy number state of each sample for a given CNP, we fitted Gaussian mixture models to the intensity distribution of all samples. (21) A copy number state was then assigned to each sample based on the posterior probability of the mixture model. We used the assigned copy number states in a logistic regression model with the following covariables: age at onset of disease, country of origin and gender. To correct for population stratification we also used as covariables the first two dimensions of the multi-dimensional scaling (MDS) analysis done with PLINK v1.06 (23) on a set of 15,000 unlinked autosomal SNPs. (9) For survival analyses we used the copy number states in a multivariate Cox regression model with age at onset, country of origin, site of onset, gender and the first two dimensions of the MDS analysis. Hardy-Weinberg equilibrium was tested with PLINK. For selection of CNPs for replication experiments, we used a threshold of p=0.01.

In the replication stage we used a similar approach: for the CNPs that were tested with qPCR, we fitted Gaussian mixtures to the intensity measure of the normalised qPCR data (as described above) and performed logistic regression analyses on the assigned copy number states with the country of origin and gender as covariates. For the CNP that was tested using fragment analysis, we used the genotypes in a Cox regression model. For calculation of LD between CNPs and neighbouring SNPs we downloaded SNP genotypes from HapMap CEU samples (r23a, http://hapmap.ncbi.nlm.nih.gov) and CNP genotypes for these samples (13) and used Haploview v1.0 (24) to compute LD between the CNPs with SNPs within 100 kb from the reported CNP start and end coordinates.

RESULTS

We used the map of CNPs as reported previously, which contains 1,292 autosomal CNPs. (13) For 387 CNPs (30%) our platform did not contain any mapping markers, and therefore could not be tested. Of the 905 remaining CNPs 652 had to be discarded because they consisted of only one component (copy number class) and therefore could not be tested for association, or because of insufficient data quality that did not allow the fitting of an appropriate model (Figure 1). The remaining 253 CNPs were tested for association with ALS susceptibility and with disease duration. The included CNPs did not encompass any known ALS-associated genes (3), and there were no CNPs in the vicinity (<250kb) of ALS associated genes *UNC13A* and *C9ORF72* (9). Probes in the SMN locus showed monomorphic intensity distributions and could therefore not be tested for association (not shown).

In the initial (genome-wide) stage we included 1,350 ALS patients and 3,262 controls from The Netherlands, Belgium and Sweden (Table 1). The results of the genome-wide stage are depicted in Figure 2. Seven CNPs showed nominal association with ALS susceptibility ($p \leq 0.01$) and were examined more closely (Table 2). To identify obvious genotyping errors we compared the observed frequencies with the allele frequencies reported in HapMap. (13) For five of the seven CNPs the allele frequencies in controls were comparable to those described in HapMap ($p>0.001$, Chi squared test). However, two of the seven (CNPs mc_10359 and mc_11963) showed marked deviations from the HapMap genotypes ($p < 2x10^{-16}$ and $p=6x10^{-8}$, Chi squared test), suggesting that these were genotyping artefacts. Indeed, these two CNPs were not in Hardy-Weinberg equilibrium (Table 2). In addition, we compared the intensity data from the Illumina array of 77 samples with qPCR data for the seven CNPs. Of three CNPs the qPCR results were in disagreement with the Illumina data (not shown). Based on comparison with HapMap allele frequencies, Hardy-Weinberg equilibrium and qPCR data, we considered three CNPs likely artefacts and we dropped them from further analyses. We selected 4 CNPs (CNPs mc_340, mc_865, mc_1179, mc_11770) for additional replication experiments. 230 Dutch patients included in the genome-wide stage were tested for the *C9ORF72* repeat expansion that has been recently reported as a cause of familial and sporadic ALS. (7, 8, 25) No data on *C9ORF72* repeat expansions of the Belgian and Swedish populations were available. To assess whether these subjects might influence the results, we repeated the association analysis after removal of known carriers of this mutation. However, removal of the 9 patients (4% of the tested subjects) who carried this mutation did not alter the results of the association analyses (not shown).

We tested the selected CNPs from the genome-wide stage in an independent population using TaqMan qPCR experiments (Methods). In total we included 2,091 patients and 1,987 controls from The Netherlands, Germany and Poland in the replication experiment (Table 1). All four CNPs were compatible with diallelic deletion polymorphisms (tables 2 and 3 and ref (13)). CNP mc_340 actually consisted of two overlapping CNPs, indicated by the different qPCR profiles generated by two adjacent Taqman probes, that each targeted a different CNP. We therefore tested these two CNPs individually (Table 3). The tested CNPs were in Hardy-Weinberg equilibrium ($p \geq 0.01$, Table 3) and the observed allele frequencies in controls in the replication experiment of all CNPs were similar to those in the genome-wide stage (except CNP 340 for which comparison between the twp phases was not possible) ($p \geq 0.01$, Chi squared test), thereby increasing confidence in the genotyping accuracy. Both CNPs failed to show significant associations in the replication population. We calculated that we had 80% power to detect a 1.5% difference in minor allele frequency given the sample size of our replication population.

Table 2: results of the discovery stage.

CNP	Position	HapMap	CN states	Freq. controls	Freq. cases	*P*	HWE *P*	Validated
mc_10359	Chr 2: 110.2 Mb	1:0.4%-2:99.2%-3:0.4%	1,2	51%:49%	54%:46%	$6.3x10^{-3}$	$5.3 x10^{-122}$	No
mc_340	Chr 2: 212.9 Mb	1:15%-2:85%	1,2,3	0.4%:15%: 85%	1%:12%:87%	$3.7 x10^{-3}$	$1.2 x10^{-1}$	Yes
mc_865	Chr 5: 140.2 Mb	0:1%-1:9%-2:90%	1,2,3	1%:18%:81%	1%:13%:86%	$8.3 x10^{-3}$	1	Yes
mc_1179	Chr 7: 142.2 Mb	0:21%-1:39%-2:40%	1,2,3	17%:48%:35%:	15%:46%:39%	$3.8 x10^{-3}$	$8.3 x10^{-1}$	Yes
mc_11770	Chr 10: 82.9 Mb	1:4%-2:95%	1,2,3	0.3%:8%:92%	0%:7%:93%	$4.9 x10^{-3}$	$1.3 x10^{-1}$	Yes
mc_11807	Chr 10: 133.5 Mb	1:8%-2:92%	1,2	7%:93%	5%:95%	$7.2 x10^{-3}$	$5.0 x10^{-2}$	No
mc_11963	Chr 12: 17.9 Mb	2:96%-3:4%	1,2	68%:32%	62%:38%	$4.8 x10^{-3}$	$8.7 x10^{-45}$	No

CNPs with a nominal p ≤ 0.01. Table shows for each CNP the genomic position, the frequency of copy-number states in the HapMap sample as reported by Mccarroll et al. 2008, the number of different fitted copy-number states (not necessarily identical to the underlying number of copies), the frequency of copy number states in controls and cases, logistic regression p value, Hardy-Weinberg equilibrium p value, validation by qPCR.

In addition to the susceptibility association analysis, we tested each CNP for association with the duration of disease (Figure 2). Three of the tested CNPs in the genome-wide stage showed association with the duration of disease at P ≤ 0.01 (Table 4). Two of these CNPs showed marked differences with the expected allele frequencies based on the HapMap samples ($p<6*10^{-5}$, Chi squared test) and showed a significant deviation from Hardy–Weinberg equilibrium. We therefore considered these false positives and discarded them from further analyses. CNP mc_10311 showed no obvious signs of genotyping artefacts (Hardy-Weinberg equilibrium: p = 1; allele frequencies comparable to those in HapMap, p=0.47. Chi squared test). We tested this CNP in the replication sample using fragment analysis (Methods). However, in this population we did not find evidence for association with duration of disease (P=0.35, Table 5).

Only one of the five CNPs that were tested in the replication stage was well tagged by a neighbouring SNP that has been typed in a previous GWAS (9): CNP mc_1179, best tagSNP: r^2=0.97. (In the GWAS this SNP also showed nominal association ($p = 1.4x10^{-3}$), corroborating our results). However, the other four CNPs were poorly tagged (best tag SNP: r^2 = 0.09, 0.38, 0.02, 0.04 for CNPs mc_340, mc_865, mc_11770 and mc_10311 respectively), illustrating the additional value of the direct CNP genotyping.

Table 3: results of the replication stage.

CNP	Position	CN states	Freq controls	Freq cases	*P*	HWE *P*
mc_340 (Hs03388917_cn)	Chr 2: 212.9 Mb	1,2,3	0.6%:14.3%:85.0%	0.5%:13.7%:85.7%	0.79	0.78
mc_340 (Hs03421659_cn)	Chr 2: 212.9 Mb	1,2,3	4.4%:33.5%:62.1%	4.0%:35.0%:61.0%	0.29	0.95
mc_865	Chr 5: 140.2 Mb	1,2,3	0.8%:19.8%:79.4%	1.0%:18.0%:81.0%	0.49	0.11
mc_1179	Chr 7: 142.2 Mb	1,2,3	15.0%:50.0%:35.0%	16.0%:47.0%:37.0%	0.86	0.01
mc_11770	Chr 10: 82.9 Mb	1,2,3	0.2%:6.2%:93.5%	0.2%:6.2%:93.5%	0.49	0.18

Table shows for each CNP the genomic position, the number of fitted copy-number states (not necessarily identical to the underlying number of copies), the frequency of each copy number state in controls and in cases, logistic regression p value, Hardy-Weinberg equilibrium p value. Please note that CNP mc_340 consisted of two overlapping CNPs, which were tested individually.

Table 4: results for the survival analyses in the genome-wide stage.

CNP	Position	HapMap	CN states	Freq. cases	HWE *P*	*P* survival	HR (95%CI)
mc_10118	Chr 1:111.6 Mb	1: 1.1% 2: 98.9%	1,2	0.83:0.17	0	$4.4x10^{-4}$	1.35 (1.14 – 1.59)
mc_10311	Chr 2:78.6 Mb	1: 1.1% 2: 98.9%	1,2	0.03:0.97	1	$3.1x10^{-4}$	1.99 (1.37 – 2.89)
mc_11052	Chr 6:35.6 Mb	2: 98.5% 3: 1.5%	1,2	0.64:0.36	$5.7x10^{-240}$	$4.5x10^{-3}$	1.21 (1.06 – 1.39)

CNPs with a nominal association with disease duration at $p \leq 0.01$. Table shows for each CNP the genomic position, the frequency of copy-number states in the HapMap sample as reported by Mccarroll et al. 2008, the number of different fitted copy-number states (not necessarily identical to the underlying number of copies), the frequency of copy number states in cases, Hardy-Weinberg equilibrium p value, Cox regression p value and hazard ratios with 95% confidence interval.

Table 5: results of the survival analysis in the replication cohort.

CNP	Freq cases	HWE *P*	*P*	HR (95%CI)
mc_10311	Del/Del: 0%; Del/Nd: 1.5%; Wt/Nd: 98.5%	1	0.35	1.33 (0.73 – 2.42)

Table shows the frequency of deletion and non-deletion alleles (designated as "Nd"), Hardy-Weinberg equilibrium p value, Cox regression p value and hazard ratio with 95% confidence interval.

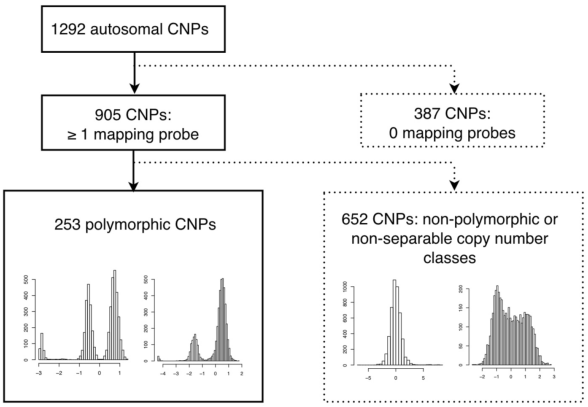


Figure 1. Study overview.

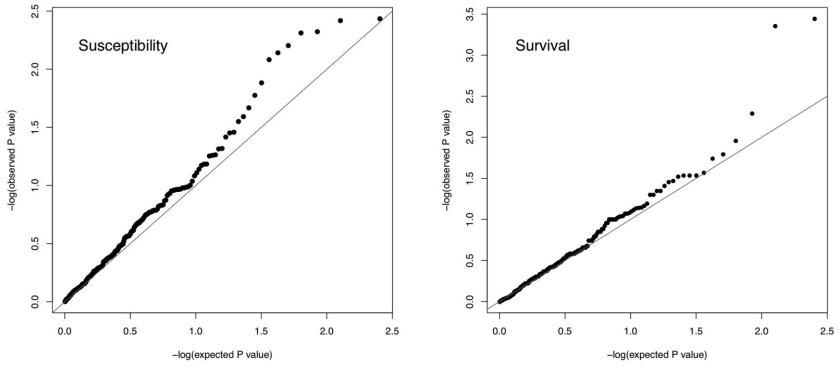


Figure 2. Quantile-quantile plots for the p values from the tested CNPs for association with ALS susceptibility (left) and duration of disease (right). = 1.13 (susceptibility); = 1.07 (survival).

DISCUSSION

In this CNP association study we examined the role of common copy-number variants as potential risk factors for ALS susceptibility and as modifiers of disease progression. We used a validated set of CNPs and tested these for association in a large, well characterised study population. Five CNPs from the genome-wide discovery stage (four associated with ALS susceptibility and one with disease progression) were selected for follow-up, but none showed evidence of significant association in replication experiments.

The role of CNPs in ALS has not been studied before in a genome-wide survey. While previous studies focused on rare, large CNVs in ALS (10-12), we investigated common (polymorphic) CNVs. Due to the different methodology we could investigate much smaller CNVs: the median size of the studied CNP in this study was 10 kb (compared for example to >150 kb in ref (10)). The spectrum of studied genetic variants in this study is therefore complementary to previous studies on rare CNVs. (10)

The systematic assessment of the role of CNPs in ALS pathogenesis is of value when considering the known disease associated CNPs and that only a part of CNPs has been indirectly assessed through linkage disequilibrium-based association testing in previous GWASs. (13, 26) (13) The latter is illustrated by the observation that only one of the most significant CNPs our study would have been captured by our previous SNP-based GWAS. (9)

We used a map of thoroughly ascertained and validated CNPs as a starting point for this study. (13) However, only a part of these CNPs could be typed on our platform: for 30% the used array did not contain any mapping probes. For a part of the remaining CNPs the intensity distribution was either of insufficient data quality or was monomorphic, i.e.: the data did not allow distinguishing between different copy number classes. The resulting set of studied CNPs is therefore relatively small. For comparison, in the Wellcome Trust Case Control Consortium CNP association study that used a dedicated CNP-targeted array, more than 3,400 CNPs were genotyped. (27) Notably, however, this was only a fraction of the ~11,000 CNP regions that were targeted by the employed array in that study; the remainder could not be genotyped. It was estimated that while more than half of these non-typeable CNPs were in fact polymorphic, it was impossible to distinguish different copy number classes, precluding association analysis. This indicates that even with a dedicated platform the genotyping of polymorphic CNVs (especially multiallelic, complex) CNVs remains challenging.

Our study, and CNP association studies in general, are hampered by several factors that are less of a problem in SNP GWAS. For example, the accurate genotyping of CNPs in a case-control study is much more difficult than SNP genotyping. Because the underlying intensity data (array probe intensities or qPCR data) are a quantitative measure, the discrimination between different copy number classes is much more error-prone than the allelic discrimination in SNP genotyping. As a result of misclassification of CNP genotypes, the statistical power to detect true associations can be reduced. (28) In addition, CNP association studies are much more prone to spurious results due to calling artefacts resulting from different sources, including differences in DNA source, DNA handling and processing (see for example references (21, 27, 29)). We have tried to eliminate these as much as possible using robust statistical analysis tools, which, given the lack of significant inflation of p values (Figure 2), appeared to work well. Most importantly, we used an independent replication experiment as a means of validation.

The fact that we did not identify associated CNPs is in agreement with the WTCCC study: in this large CNP association study in eight common complex disease traits, no novel associations were found in addition to loci that had already been identified through previous GWASs. (27) It is estimated that perhaps less than 5% of identified GWAS hits

are explained by SNP-tagged CNPs. (30) This leads to the notion that the contribution of CNPs to the genetic basis of disease is probably modest, and does not explain the a large part of "missing heritability" of most complex traits. (30) However, given the observation that many CNPs have not been captured in our current study or by previous GWASs, the systematic characterisation of CNPs may yield new insights into the genetic underpinnings of ALS.

In conclusion, our study did not identify ALS-associated CNPs. Only a part of known CNPs could be interrogated using our methodology, so although many CNPs have probably been indirectly captured in previous GWASs through tagging SNPs, a large number remains unexplored. The accurate typing of CNPs remains a challenge, for which future studies may provide a definitive solution.

REFERENCES

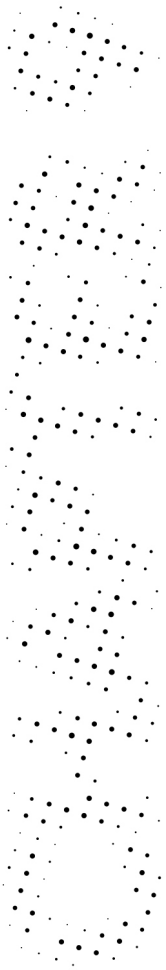
1. Kiernan, M.C., Vucic, S., Cheah, B.C., Turner, M.R., Eisen, A., Hardiman, O., Burrell, J.R. and Zoing, M.C. (2011) Amyotrophic lateral sclerosis. *Lancet*, 377, 942-955.
2. Miller, R.G., Mitchell, J.D., Lyon, M. and Moore, D.H. (2007) Riluzole for amyotrophic lateral sclerosis (ALS)/ motor neuron disease (MND). *Cochrane Database Syst Rev*, CD001447.
3. Andersen, P.M. and Al-Chalabi, A. (2011) Clinical genetics of amyotrophic lateral sclerosis: what do we really know? *Nat Rev Neurol*, 7, 603-615.
4. Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M. *et al.* (2012) The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur J Hum Genet*, Epub ahead of print.
5. Al-Chalabi, A., Fang, F., Hanby, M.F., Leigh, P.N., Shaw, C.E., Ye, W. and Rijsdijk, F. (2010) An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry*, 81, 1324-1326.
6. Elden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M. *et al.* (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, 466, 1069-1075.
7. Dejesus-Hernandez, M., Mackenzie, Ian R., Boeve, Bradley F., Boxer, Adam L., Baker, M., Rutherford, Nicola J., Nicholson, Alexandra M., Finch, NiCole A., Flynn, H., Adamson, J. *et al.* (2011) Expanded GGGGCC Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron*, 1-13.
8. Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., Van Swieten, J.C., Myllykangas, L. *et al.* (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, 72, 257-268.
9. Van Es, M.A., Veldink, J.H., Saris, C.G.J., Blauw, H.M., Van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B. *et al.* (2009) Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet*, 41, 1083-1087.
10. Blauw, H.M., Al-Chalabi, A., Andersen, P.M., van Vught, P.W., Diekstra, F.P., van Es, M.A., Saris, C.G., Groen, E.J., van Rheenen, W., Koppers, M. *et al.* (2010) A large genome scan for rare CNVs in amyotrophic lateral sclerosis. *Hum Mol Genet*, 19, 4091-4099.
11. Wain, L.V., Pedroso, I., Landers, J.E., Breen, G., Shaw, C.E., Leigh, P.N., Brown, R.H., Tobin, M.D. and Al-Chalabi, A. (2009) The role of copy number variation in susceptibility to amyotrophic lateral sclerosis: genome-wide association study and comparison with published loci. *PLoS One*, 4, e8175.
12. Cronin, S., Blauw, H.M., Veldink, J.H., van Es, M.A., Ophoff, R.A., Bradley, D.G., van den Berg, L.H. and Hardiman, O. (2008) Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. *Hum Mol Genet*, 17, 3392-3398.
13. McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy

number variation. *Nat Genet*, 40, 1166-1174.
14. Hollox, E.J., Huffmeier, U., Zeeuwen, P.L., Palla, R., Lascorz, J., Rodijk-Olthuis, D., van de Kerkhof, P.C., Traupe, H., de Jongh, G., den Heijer, M. *et al.* (2008) Psoriasis is associated with increased beta-defensin genomic copy number. *Nat Genet*, 40, 23-25.
15. de Cid, R., Riveira-Munoz, E., Zeeuwen, P.L., Robarge, J., Liao, W., Dannhauser, E.N., Giardina, E., Stuart, P.E., Nair, R., Helms, C. *et al.* (2009) Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat Genet*, 41, 211-215.
16. Aitman, T.J., Dong, R., Vyse, T.J., Norsworthy, P.J., Johnson, M.D., Smith, J., Mangion, J., Roberton-Lowe, C., Marshall, A.J., Petretto, E. *et al.* (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. *Nature*, 439, 851-855.
17. McCarroll, S.A., Huett, A., Kuballa, P., Chilewski, S.D., Landry, A., Goyette, P., Zody, M.C., Hall, J.L., Brant, S.R., Cho, J.H. *et al.* (2008) Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat Genet*, 40, 1107-1112.
18. Willer, C.J., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C. *et al.* (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet*, 41, 25-34.
19. Blauw, H.M., Barnes, C.P., van Vught, P.W., van Rheenen, W., Verheul, M., Cuppen, E., Veldink, J.H. and van den Berg, L.H. (2012) SMN1 gene duplications are associated with sporadic ALS. *Neurology*, 78, 776-780.
20. Simon-Sanchez, J., Schulte, C., Bras, J.M., Sharma, M., Gibbs, J.R., Berg, D., Paisan-Ruiz, C., Lichtner, P., Scholz, S.W., Hernandez, D.G. *et al.* (2009) Genome-wide association study reveals genetic risk underlying Parkinson's disease. *Nat Genet*, 41, 1308-1312.
21. Barnes, C., Plagnol, V., Fitzgerald, T., Redon, R., Marchini, J., Clayton, D. and Hurles, M.E. (2008) A robust statistical method for case-control association testing with copy number variation. *Nat Genet*, 40, 1245-1252.
22. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. and Vandesompele, J. (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol*, 8, R19.
23. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81, 559-575.
24. Barrett, J.C., Fry, B., Maller, J. and Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, 21, 263-265.
25. van Rheenen, W., van Blitterswijk, M., Huisman, M.H., Vlam, L., van Doormaal, P.T., Seelen, M., Medic, J., Dooijes, D., de Visser, M., van der Kooi, A.J. *et al.* (2012) Hexanucleotide repeat expansions in C9ORF72 in the spectrum of motor neuron diseases. *Neurology*, 79, 878-882.
26. Campbell, C.D., Sampas, N., Tsalenko, A., Sudmant, P.H., Kidd, J.M., Malig, M., Vu, T.H., Vives, L., Tsang, P., Bruhn, L. *et al.* (2011) Population-genetic properties of differentiated human copy-number polymorphisms. *Am J Hum Genet*, 88, 317-332.
27. Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E. *et al.* (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 464, 713-720.
28. Kim, W., Gordon, D., Sebat, J., Ye, K.Q. and Finch, S.J. (2008) Computing power and sample size for case-control association studies with copy number polymorphism: application of mixture-based likelihood ratio test. *PLoS One*, 3, e3475.
29. Aldhous, M.C., Abu Bakar, S., Prescott, N.J., Palla, R., Soo, K., Mansfield, J.C., Mathew, C.G., Satsangi, J. and Armour, J.A.L. (2010) Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Hum Mol Genet*, 19, 4930-4938.
30. Conrad, D.F., Pinto, D., Redon, R., Feuk, L., Gokcumen, O., Zhang, Y., Aerts, J., Andrews, T.D., Barnes, C., Campbell, P. *et al.* (2009) Origins and functional impact of copy number variation in the human genome. *Nature*, 464, 704-712.

SUPPLEMENTARY MATERIAL

Supplementary Table 1: used TaqMan assays for the replication stage.

ASSAY	CHR	POSITION	CNP ID
Hs03388917_cn	2	212895000	340
Hs03421659_cn	2	212895693	340
Hs00800226_cn	5	140207609	865
Hs01415729_cn	5	140207698	865
Hs03114440_cn	7	142162984	1179
Hs03166397_cn	7	142165809	1179
Hs04376951_cn	10	82872598	11770
Hs03747412_cn	10	82874224	11770
Hs03754977_cn	10	133498307	11807
Hs03750569_cn	10	133503725	11807



NIPA1 POLYALANINE REPEAT EXPANSIONS ARE ASSOCIATED WITH AMYOTROPHIC LATERAL SCLEROSIS

Human Molecular Genetics (2012)

Hylke M. Blauw, MD,[1] Wouter van Rheenen, MD,[1] Max Koppers, MSc,[1] Philip Van Damme, MD, PhD,[2] Stefan Waibel, MD, PhD,[3] Robin Lemmens, MD, PhD,[2] Paul W. J. van Vught PhD,[1] Thomas Meyer, MD, PhD,[4] Claudia Schulte, PhD,[5] Thomas Gasser, MD, PhD,[5] Edwin Cuppen, PhD,[6] R. Jeroen Pasterkamp, PhD,[7] Wim Robberecht,[2] Albert C. Ludolph,[3] Jan H. Veldink, MD, PhD, [1] Leonard H. van den Berg, MD, PhD, [1,*]

[1]Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands; [2]Department of Neurology, University Hospital Leuven, University of Leuven, Leuven, Belgium and Vesalius Research Center, VIB, Leuven, Belgium; [3]Department of Neurology, University of Ulm, Ulm, Germany; [4]Department of Neurology, Charité University Hospital, Humboldt-University, Berlin, Germany; [5]Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and German Center for Neurodegenerative Diseases, Tübingen, Germany; [6]Hubrecht Institute for Developmental Biology and Stem Cell Research, Royal Netherlands Academy of Sciences and University Medical Center Utrecht, The Netherlands. [7]Department of Neuroloscience and Pharmacology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Utrecht, The Netherlands.

The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint first authors, and the last two authors should be regarded as joint last authors. *To whom correspondence should be addressed: Department of Neurology, UMC Utrecht, Heidelberglaan 100, 3584CX, Utrecht, The Netherlands; Telephone: 0031 88 7555555; E-mail: l.h.vandenberg@umcutrecht.nl

ABSTRACT

Deletions of *NIPA1* have been associated with a higher susceptibility to amyotrophic lateral sclerosis (ALS). The exact role of genetic variation in *NIPA1* in ALS susceptibility and disease course is, however, not known. We sequenced the entire coding sequence of *NIPA1* and genotyped a polyalanine repeat located in the first exon of *NIPA1*. A total of 2292 ALS patients and 2777 controls from three independent European populations were included. We identified two sequence variants that have a potentially damaging effect on *NIPA1* protein function. Both variants were identified in ALS patients; no damaging variants were found in controls. Secondly, we found a significant effect of "long" polyalanine repeat alleles on disease susceptibility: odds ratio = 1.71, $p = 1.6 \times 10^{-4}$. Our analyses also revealed a significant effect of "long" alleles on patient survival (hazard ratio (HR) = 1.60, $p = 4.2 \times 10^{-4}$) and on the age at onset of symptoms (HR = 1.37, $p = 4.6 \times 10^{-3}$). In patients carrying "long" alleles, median survival was three months shorter than patients with "normal" genotypes and onset of symptoms occurred 3.6 years earlier. Our data show that *NIPA1* polyalanine repeat expansions are a common risk factor for ALS and modulate disease course.

INTRODUCTION

Amyotrophic lateral sclerosis is an invariably fatal motor neuron disease. It is characterized by the selective demise of motor neurons in the brain and the spinal cord, leading to spasticity and muscle weakness. To date, there is no curative therapy and patients usually succumb to respiratory failure within five years after the onset of symptoms. (1) The majority of patients has no family history of the disease and are said to be sporadic. Although there is increasing evidence that genetic risk factors contribute to pathology in sporadic ALS (2-4), few well-established susceptibility genes have been identified. Likewise, little is known about genetic factors that determine the age at onset of the disease and the rate of disease progression.

Recently, we reported results from a genome-wide study, suggesting that rare deletions of *NIPA1* (non-imprinted in Prader-Willi/ Angelman syndrome *1*, MIM ID 608145) increase the risk of ALS. (5) Mutations in *NIPA1* are known to cause Hereditary Spastic Paraplegia (HSP) type 6, a neurodegenerative disease characterized by an (upper) motor neuron phenotype, and cause progressive paralysis in animal models. (6, 7) *NIPA1* is, therefore, a plausible candidate gene from a genetic and biological point of view.

The 5' end of *NIPA1* contains a polyalanine repeat of 12-13 alanine residues. The majority of this repeat is encoded by a polymorphic (GCG)n repeat, with GCG7 and GCG8 being the most frequent alleles (allele frequencies of 0.20 and 0.78 respectively). (8) Several disease phenotypes are caused by polyalanine expansions in other genes, such as oculopharyngeal muscular dystrophy (OPMD) and Ondine syndrome. (9) Recently, repeat expansions in *C9ORF72* and *ATXN2* have been described as important risk factors for ALS. (2, 10-13) The effect of *NIPA1* polyalanine repeat expansions or contractions has, however, not been studied before.

To characterize the effects of genetic variation in *NIPA1* on ALS pathogenesis we carried out a large genetic association study in three European populations. We systematically sequenced all coding sequences and genotyped the polyalanine repeat in the first exon using an in-house developed assay.

RESULTS

Our analyses included 2292 ALS patients and 2777 unaffected controls from The Netherlands, Germany and Belgium (Table 1).

In the mutation screen we identified twelve unique, non-synonymous sequence variants in thirteen individuals: seven in sporadic ALS patients and six in controls (Table 2). Ten of twelve are predicted to be benign variants, while two are predicted to be damaging. The latter were identified exclusively in patients. While variants (both damaging and benign) in *NIPA1* are more prevalent in ALS patients compared to controls, this did not reach statistical significance ($p = 0.20$ for damaging variants, $p = 0.36$ for non-synonymous variants, one-sided Fisher exact test). We did not find mutations in 70 familial cases. The patients with damaging mutations did have an upper motor-predominant phenotype.

We then genotyped a polymorphic GCG repeat in our study population, which is located in the first exon of *NIPA1*. Consistent with previous reports (8), the two most frequent alleles in the control population are (GCG)7 and (GCG)8, with allele frequencies of 0.22 and 0.75, respectively (Table S1). Shorter and longer repeat lengths are very rare in controls, with (GCG)10 being the most prevalent with an allele frequency of about 1%. To determine whether the polyalanine repeat length was associated with ALS disease status, we first tested if the allele frequencies differed between patients and controls. The allele frequency differed significantly between patients and controls ($p = 3.6 \times 10^{-4}$, Cochran-Mantel-Haenszel test), without heterogeneity across the different populations (Woolf test $p = 0.75$).

To evaluate the effect of short or long repeat length, we trichotomized individuals as either "normal" ((GCG)7 or (GCG)8), "short" (carriers of alleles containing < 7 GCG repeats), or "long" (carriers with alleles containing > 8 GCG repeats. Logistic regression analysis revealed a significant effect of "long" GCG repeat length on susceptibility: $p = 1.6 \times 10^{-4}$, OR = 1.71, 95% CI = 1.30 – 2.26 (Table 3). This effect was consistent throughout the three analyzed populations (OR = 1.68, 2.09 and 1.22 for the Dutch, German and Belgian populations, respectively). The effect of short alleles was not significant.

We then examined the effect of the GCG repeat length on survival and on age at onset of the disease. For the survival analyses, data was available for Dutch and Belgian populations. Carriers of a long GCG repeat had shorter median survival (29.8 months) compared to carriers of normal (32.8 months) GCG repeats ($p = 4.2 \times 10^{-4}$, hazard ratio (HR) = 1.60, 95% CI = 1.23– 2.06; Table 4, Figure 1A, Figure S1). There was no clear effect of repeat-length differences on survival in the Belgian population. Because there was evidence of some non-proportionality in the survival analyses with regard to the covariable age at onset only, we additionally analyzed survival using a non-parametric analysis of the effect of *NIPA1* polyalanine repeat length on survival using weighted Cox regression. This produced similar results: a significant effect of "long" alleles on the disease duration ($p = 2.5 \times 10^{-5}$, HR = 1.63, 95% CI = 1.30 – 2.04).

Our analyses also revealed a significant effect of the GCG repeat length on the age at onset of symptoms in ALS patients. The median age at onset of the disease was 58.3 years for carriers of a long GCG repeat, compared to 61.9 years for carriers of normal alleles ($p = 4.6 \times 10^{-3}$, HR = 1.37, 95% CI = 1.10 – 1.71; Table 5, Figure 1B, Figure S2). This effect was present in all three populations: HR = 1.30, 1.74 and 1.42 in Dutch, German and Belgian populations, respectively).

Table 1. Summary of the study populations.

	Subjects, n	Mean age (years, +/- SD)	Female, %
The Netherlands			
ALS	924	62.3 (+/- 11.7)	40
Controls	1729	62.1 (+/- 11.5)	47
Germany			
ALS	1006	59.8 (+/- 11.9)	42
Controls	632	59.2 (+/- 7.2)	40
Belgium			
ALS	362	58.6 (+/- 12.9)	41
Controls	416	57.7 (+/- 14.4)	57
Total			
ALS	2292	60.7 (+/- 12.0)	41
Controls	2777	60.8 (+/- 11.3)	47

SD = standard deviation.

Table 2. Identified sequence variants.

Phenotype	AA change	Functional effect
ALS	I81T	Possibly damaging
ALS	A86G	Benign
ALS	I120M	Benign
ALS	V162M	Benign
ALS	M189I	Benign
ALS	P221L	Probably damaging
ALS	R281Q	Benign
Control	G2E	Benign
Control	A11V	Benign
Control	A12V	Benign
Control	A86G	Benign
Control	V301I	Benign
Control	V303L	Benign

Table 3. *NIPA1* polyalanine alleles in ALS patients and controls.

	Controls (%)	ALS (%)	OR (+/- 95% CI)	p
The Netherlands				
Short	7 (0.4)	10 (1.1)	2.44 (0.90 - 6.59)	0.08
Normal	1655 (95.7)	855 (92.5)		
Long	67 (3.9)	59 (6.4)	1.68 (1.17 - 2.41)	4.9×10^{-3}
Total	1729 (100.0)	924 (100.0)		
Germany				
Short	2 (0.3)	5 (0.5)	1.65 (0.32 - 8.55)	0.55
Normal	613 (97.0)	951 (94.5)		
Long	17 (2.7)	50 (5.0)	2.09 (1.18 - 3.72)	0.01
Total	632 (100.0)	1006 (100.0)		
Belgium				
Short	4 (1.0)	1 (0.3)	0.32 (0.03 - 3.18)	0.33
Normal	397 (95.4)	345 (95.3)		
Long	15 (3.6)	16 (4.4)	1.22 (0.59 - 2.53)	0.59
Total	416 (100.0)	362 (100.0)		
Combined				
Short	13 (0.5)	16 (0.7)	1.72 (0.79 - 3.75)	0.17
Normal	2665 (96.0)	2151 (93.8)		
Long	99 (3.6)	125 (5.5)	1.71 (1.3 - 2.26)	1.6×10^{-4}
Total	2777 (100.0)	2292 (100.0)		

OR = odds ratio. CI = confidence interval. Normal: carriers of $(GCG)_7$ or $(GCG)_8$; Short: carriers of alleles containing < 7 GCG repeats; Long: carriers with alleles containing > 8 GCG repeats.

Table 4: Patient survival.

	N	Median survival (months, SD)	HR (+/- 95% CI)	p
The Netherlands				
Short	10	52.7 (13.0)	0.67 (0.32 - 1.41)	0.29
Normal	855	33.3 (40.1)		
Long	59	27.4 (21.4)	1.73 (1.30 - 2.32)	$2.1x10^{-4}$
Belgium				
Short	1	24.0 (NA)	2.75 (0.38 - 19.77)	0.32
Normal	345	30.0 (28.4)		
Long	16	35.8 (18.0)	1.16 (0.66 - 2.06)	0.60
Combined				
Short	11	47.4 (14.6)	0.75 (0.37 - 1.51)	0.42
Normal	1200	32.8 (37.2)		
Long	75	29.8 (20.7)	1.59 (1.23 - 2.06)	$4.2x10^{-4}$

SD = standard deviation. HR = hazard ratio. CI = confidence interval. Normal: carriers of (GCG)7 or (GCG)8; Short: carriers of alleles containing < 7 GCG repeats; Long: carriers with alleles containing > 8 GCG repeats.

Table 5: Age at onset of disease.

	N	Median age at onset (years, SD)	HR (+/- 95% CI)	p
The Netherlands				
Short	10	62.9 (13.2)	1.02 (0.53 - 1.97)	0.95
Normal	855	61.8 (11.9)		
Long	59	58.3 (10.7)	1.30 (1.00 - 1.69)	0.05
Germany				
Short	5	67.4 (11.1)	0.38 (0.05 - 2.69)	0.33
Normal	951	62.3 (11.9)		
Long	50	58.7 (11.7)	1.74 (0.94 - 3.23)	0.08
Belgium				
Short	1	64.0 (NA)	0.84 (0.12 - 6.02)	0.86
Normal	345	61.0 (12.6)		
Long	16	51.0 (14.2)	1.42 (0.85 - 2.35)	0.18
Combined				
Short	16	64.0 (11.9)	0.88 (0.49 - 1.60)	0.68
Normal	2151	61.9 (12.0)		
Long	125	58.3 (11.6)	1.37 (1.10 - 1.71)	$4.6x10^{-3}$

SD = standard deviation. HR = hazard ratio. CI = confidence interval. Normal: carriers of $(GCG)_7$ or $(GCG)_8$; Short: carriers of alleles containing < 7 GCG repeats; Long: carriers with alleles containing > 8 GCG repeats.

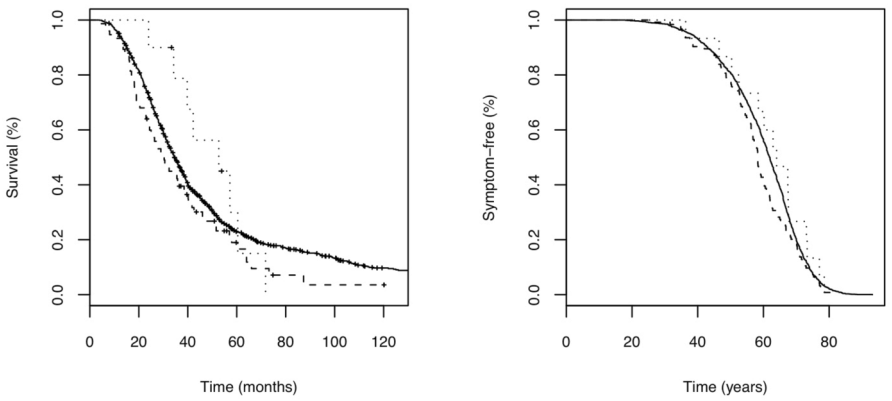


Figure 1. Kaplan-Meier plots showing the survival (A) and age at onset for "normal" (uninterrupted line), "short" (dotted line) and "long" (dashed line) NIPA1 polyalanine repeat alleles.

DISCUSSION

Our data show that *NIPA1* alleles with a long polyalanine repeat length confer an increased risk of ALS, are associated with short survival and, independently, with a younger age at onset of the disease. Together with previous data from a genome-wide scan for rare copy number variant which suggested that rare deletions of *NIPA1* are associated with ALS (5), our current data add support to *NIPA1* being a modulator of susceptibility and disease course in ALS. This is the first example of a genetic variant that increases the susceptibility to the disease as well as modulating the disease onset and survival.

We tested three independent European study populations for susceptibility, patient survival and age at onset of disease. The effects of *NIPA1* repeat-length on susceptibility and on the age at onset of symptoms show similar effects in the three tested populations. The effect on survival, however, is determined mainly by the Dutch population, possibly explained by the modest sample size of the Belgian population. The overall effect appears to be robust, but further studies will have to show whether the pathogenic effect of *NIPA1* polyalanine expansions on disease duration also exist in other populations.

Mutations in *NIPA1* are known to cause motor neuron phenotypes in humans as well as in animals. (6, 7) Polyalanine expansions, however, in *NIPA1* have not previously been associated to disease and their biological effects are unknown. In vitro studies have shown that *NIPA1* is an inhibitor of Bone Morphogenetic Protein (BMP) signaling. *NIPA1* interacts with the type II BMP receptor (BMPRII) and promotes its degradation through endocytosis and lysosomal degradation. HSP-associated mutants of *NIPA1* are less efficient at promoting BMPRII degradation than wild-type *NIPA1.* (14) Regulation of BMP signaling by NIPA1 in *Drosophila* is critical for the regulation of synaptic growth and axonal microtubules. (15) Thus, loss-of-function *NIPA1* mutations or *NIPA1* polyalanine expansions may result in defects in synapse and axon development.

Polyalanine peptides undergo various levels of conformational transition from monomeric alpha-helix to macromolecular beta sheets, which is dependent on the repeat-length. (9) Polyalanine peptides longer than 15 residues are completely converted to a beta sheet. (16) In our study we found long, but not short polyalanine repeat lengths to be harmful, and short polyalanine repeats even showed a trend towards milder phenotypes. Therefore, the mechanism of increased propensity of protein aggregation is compatible with our results.

Taken together with recent findings that intronic hexanucleotide expansions in *C9ORF72* and polyglutamine repeat expansions in *ATXN2* are associated with ALS (2, 10, 11), *NIPA1* is the third gene harboring an ALS-associated repeat expansion. It is unclear whether the biological functions of the affected genes or the repeat expansions per se determine the effect on ALS risk. A recent screening of several polyglutamine repeat-containing genes did not show associations with ALS (17), indicating that the biology of the affected genes, rather than the presence of repeat expansions per se determine the effect on ALS risk. However, the recent discovery of repeat expansions in *C9ORF72* , a non-coding region, may suggest otherwise. The question remains how many other ALS-associated repeats remain to be discovered, since this type of variation is difficult to capture even with next-generation sequencing techniques.

Our mutation screen identified two heterozygous mutations in ALS patients with a potentially damaging effect on *NIPA1* protein function, while no damaging mutations were found in controls. The two identified mutations did not include the known HSP-associated mutations T45R (6), G106R (18, 19) and A100T (20)). While we cannot exclude the possibility that *NIPA1* mutations may be a risk factor for ALS, the rarity of these mutations precludes meaningful statistical analyses. Mutation analyses in other study populations

may be useful in order to reveal the true meaning of these mutations in ALS pathogenesis.

In conclusion, our data support an important role of *NIPA1* in ALS pathogenesis and as a modulator of disease progression. Our findings point to the BMP signaling cascade as a new target for follow-up studies and may ultimately lead to the development of new treatment strategies.

MATERIALS AND METHODS

Subjects

For this study we included patients and controls from The Netherlands, Germany and Belgium. These study populations have been described in detail elsewhere. (5) In short, Dutch patients and controls were recruited in the national referral center for motor neuron diseases in the outpatient clinic of the University Hospital Utrecht, and as part of a nation-wide prospective study on motor neuron diseases in The Netherlands. German participants were recruited in the Departments of Neurology of the University Hospital in Ulm, Charité University Hospital in Berlin and University Hospital in Tuebingen. Belgian participants were recruited in the Department of Neurology of the University Hospital Gasthuisberg in Leuven. In addition, we analyzed 70 Dutch patients from 62 families with familial ALS, without known mutations in S*OD1*, *VAPB*, *ANG*, *FUS*, *TARDBP*, *VCP*, *OPTN* and *CHMP2B.* The relevant medical ethical committees approved all procedures and all participants gave written informed consent.

Sequencing and fragment-length analysis

Sequencing was performed using modifications of protocols described elsewhere.(21, 22) PCR primers were designed to cover all exonic and flanking sequences in five amplicons (Ensembl transcript ID ENST00000337435, Table S2). PCR set-ups had to be optimized for each amplicon and are available upon request. All amplicons except exon 2 were amplified using Long PCR Enzyme Mix (Fermentas, Germany). The PCR products from these reactions were used as template for sequencing reactions using BigDye chemistry (v3.1, Applied Biosystems, USA). The PCR product from exon 1 was purified and diluted before sequencing. Purified sequence product was then analyzed on 96-well 3730XL capillary sequencers (Applied Biosystems). Sequence data was analyzed for polymorphic positions using PolyPhred software for further analysis.(23) We disregarded known single-nucleotide polymorphisms (dbSNP build 129) and silent mutations in our analyses. All identified sequence variants were confirmed by independent PCR and sequencing reactions. The functional impact of identified variants was predicted using PolyPhen software (http://genetics.bwh.harvard.edu/pph/).

Genotyping of the GCG repeat in the first exon was done using fragment analysis with fluorescent-labeled primers. The region containing the GCG-repeat was amplified in a PCR reaction on genomic DNA, using a 6FAM-labeled fluorescent forward primer. The amplicon was PCR-amplified in a reaction mixture of 10 µl containing 0.4 µl long-range Taq polymerase (Long PCR Enzyme Mix, Fermentas, Germany), 200 nM dNTPs, 100 nM of forward and reverse primer, 5% dimethylsulfoxide, 1 µl PCR buffer with MgCl2 (Long PCR Enzyme Mix, Fermentas, Germany), and 50 ng of genomic DNA. PCR reaction conditions were as follows: 4 min. initial denaturation at 95˚C; 35 cycles of 20 s 94˚C, 30 s 55˚C, 4 min 68˚C; 10 min 68˚C. The PCR products were diluted 1:90 in milli-Q and transferred to a formamide solution containing a GS500-ROX ladder. Fragment analysis was performed on an ABI 3730 automated sequencer (Applied Biosystems, USA). Finally, fragment length was determined with GeneMapper version 3.7 (Applied

Biosystems, USA).

To determine the performance of our assay, we compared repeat-length data acquired with our fragment analysis assay with sequence data of 738 individuals (1476 alleles). 1444 of 1476 alleles had identical repeat lengths (97.8% agreement). In addition, all individuals with alleles other than the most common alleles ((GCG)7 and (GCG)8) were analyzed in repeated experiments by sequencing and with fragment analysis.

Statistical analyses

All statistical procedures were carried out in R 2.10.1 (http://www.r-project.org). For association analyses Fisher's exact test, Cochran-Mantel-Haenszel test (including the Woolf test, to test for possible heterogeneity among the included populations), and logistic regression were applied. In the logistic regression analyses, the effect of the polyalanine repeat length on the disease status was tested with gender, age at onset and country of origin (for combined analyses) as covariates. Effect on age at onset of the disease and the duration of the disease from onset of symptoms to death, was tested using multivariate Cox regression, with gender, site of onset (spinal or bulbar), country of origin and (for survival analyses) age at onset as covariates. Additionally, after checking the proportional hazards assumption through the scaled Schoenfeld residuals, we tested the effect on disease duration using weighted Cox regression with the "coxphw" package version 1.3.

ACKNOWLEDGEMENTS

We would like to thank all patients and healthy volunteers participating in this study. This project was supported by the Prinses Beatrix Fonds, VSB fonds, H. Kersten and M. Kersten (Kersten Foundation), The Netherlands ALS Foundation, 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. P.V.D. holds a clinical investigatorship from the FWO-Vlaanderen. R.L. is supported through research Funds of the K.U. Leuven. The work leading to this invention has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under the Health Cooperation Programme.

REFERENCES

1. Pasinelli, P. and Brown, R.H. (2006) Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat. Rev. Neurosci.*, 7, 710-723.
2. Elden, A.C., Kim, H.J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M. *et al.* (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, 466, 1069-1075.
3. Al-Chalabi, A., Fang, F., Hanby, M.F., Leigh, P.N., Shaw, C.E., Ye, W. and Rijsdijk, F. (2010) An estimate of amyotrophic lateral sclerosis heritability using twin data. *J. Neurol. Neurosurg. Psychiatry*, 81, 1324-1326.
4. van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H. *et al.* (2009) Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.*, 41, 1083-1087.
5. Blauw, H.M., Al-Chalabi, A., Andersen, P.M., van Vught, P.W., Diekstra, F.P., van Es, M.A., Saris, C.G., Groen, E.J., van Rheenen, W., Koppers, M. *et al.* (2010) A large genome scan for rare CNVs in amyotrophic lateral sclerosis. *Hum. Mol. Genet.*, 19, 4091-4099.
6. Rainier, S., Chai, J.-H., Tokarz, D., Nicholls, R.D. and Fink, J.K. (2003) NIPA1 gene mutations cause autosomal

dominant hereditary spastic paraplegia (SPG6). *Am. J. Hum. Genet.*, 73, 967-971.
7. Zhao, J., Matthies, D.S., Botzolakis, E.J., Macdonald, R.L., Blakely, R.D. and Hedera, P. (2008) Hereditary Spastic Paraplegia-Associated Mutations in the NIPA1 Gene and Its Caenorhabditis elegans Homolog Trigger Neural Degeneration In Vitro and In Vivo through a Gain-of-Function Mechanism. *J. Neurosci.*, 28, 13938-13951.
8. Chai, J.-H., Locke, D.P., Greally, J.M., Knoll, J.H.M., Ohta, T., Dunai, J., Yavor, A., Eichler, E.E. and Nicholls, R.D. (2003) Identification of four highly conserved genes between breakpoint hotspots BP1 and BP2 of the Prader-Willi/Angelman syndromes deletion region that have undergone evolutionary transposition mediated by flanking duplicons. *Am. J. Hum. Genet.*, 73, 898-925.
9. Messaed, C. and Rouleau, G.A. (2009) Molecular mechanisms underlying polyalanine diseases. *Neurobiology of Disease*, 34, 397-405.
10. DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J. *et al.* (2011) Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron*.
11. Renton, A.E., Majounie, E., Waite, A., SimÛn-S·nchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van†Swieten, J.C., Myllykangas, L. *et al.* (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*.
12. Lee, T., Li, Y.R., Ingre, C., Weber, M., Grehl, T., Gredal, O., de Carvalho, M., Meyer, T., Tysnes, O.B., Auburger, G. *et al.* (2011) Ataxin-2 intermediate-length polyglutamine expansions in European ALS patients. *Hum. Mol. Genet.*, 20, 1697-1700.
13. Van Damme, P., Veldink, J.H., van Blitterswijk, M., Corveleyn, A., van Vught, P.W., Thijs, V., Dubois, B., Matthijs, G., van den Berg, L.H. and Robberecht, W. (2011) Expanded ATXN2 CAG repeat size in ALS identifies genetic overlap between ALS and SCA2. *Neurology*, 76, 2066-2072.
14. Tsang, H.T.H., Edwards, T.L., Wang, X., Connell, J.W., Davies, R.J., Durrington, H.J., O'Kane, C.J., Luzio, J.P. and Reid, E. (2009) The hereditary spastic paraplegia proteins NIPA1, spastin and spartin are inhibitors of mammalian BMP signalling. *Hum. Mol. Genet.*, 18, 3805-3821.
15. Wang, X., Shaw, W.R., Tsang, H.T., Reid, E. and O'Kane, C.J. (2007) Drosophila spichthyin inhibits BMP signaling and regulates synaptic growth and axonal microtubules. *Nat. Neurosci.*, 10, 177-185.
16. Shinchuk, L.M., Sharma, D., Blondelle, S.E., Reixach, N., Inouye, H. and Kirschner, D.A. (2005) Poly-(L-alanine) expansions form core beta-sheets that nucleate amyloid assembly. *Proteins*, 61, 579-589.
17. Lee, T., Li, Y.R., Chesi, A., Hart, M.P., Ramos, D., Jethava, N., Hosangadi, D., Epstein, J., Hodges, B., Bonini, N.M. *et al.* (2011) Evaluating the prevalence of polyglutamine repeat expansions in amyotrophic lateral sclerosis. *Neurology*, 76, 2062-2065.
18. Chen, S., Song, C., Guo, H., Xu, P., Huang, W., Zhou, Y., Sun, J., Li, C.-X., Du, Y., Li, X. *et al.* (2005) Distinct novel mutations affecting the same base in the NIPA1 gene cause autosomal dominant hereditary spastic paraplegia in two Chinese families. *Hum Mutat*, 25, 135-141.
19. Reed, J.A., Wilkinson, P.A., Patel, H., Simpson, M.A., Chatonnet, A., Robay, D., Patton, M.A., Crosby, A.H. and Warner, T.T. (2005) A novel NIPA1 mutation associated with a pure form of autosomal dominant hereditary spastic paraplegia. *Neurogenetics*, 6, 79-84.
20. Kaneko, S., Kawarai, T., Yip, E., Salehi-Rad, S., Sato, C., Orlacchio, A., Bernardi, G., Liang, Y., Hasegawa, H., Rogaeva, E. *et al.* (2006) Novel SPG6 mutation p.A100T in a Japanese family with autosomal dominant form of hereditary spastic paraplegia. *Mov Disord*, 21, 1531-1533.
21. van Boxtel, R., Toonen, P.W., Verheul, M., van Roekel, H.S., Nijman, I.J., Guryev, V. and Cuppen, E. (2008) Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. *BMC Genomics*, 9, 460.
22. Smits, B.M., Mudde, J.B., van de Belt, J., Verheul, M., Olivier, J., Homberg, J., Guryev, V., Cools, A.R., Ellenbroek, B.A., Plasterk, R.H. *et al.* (2006) Generation of gene knockouts and mutant models in the laboratory rat by ENU-driven target-selected mutagenesis. *Pharmacogenet Genomics*, 16, 159-169.
23. Nickerson, D.A., Tobe, V.O. and Taylor, S.L. (1997) PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. *Nucleic Acids Res*, 25, 2745-2751.

SUPPLEMENTARY MATERIAL

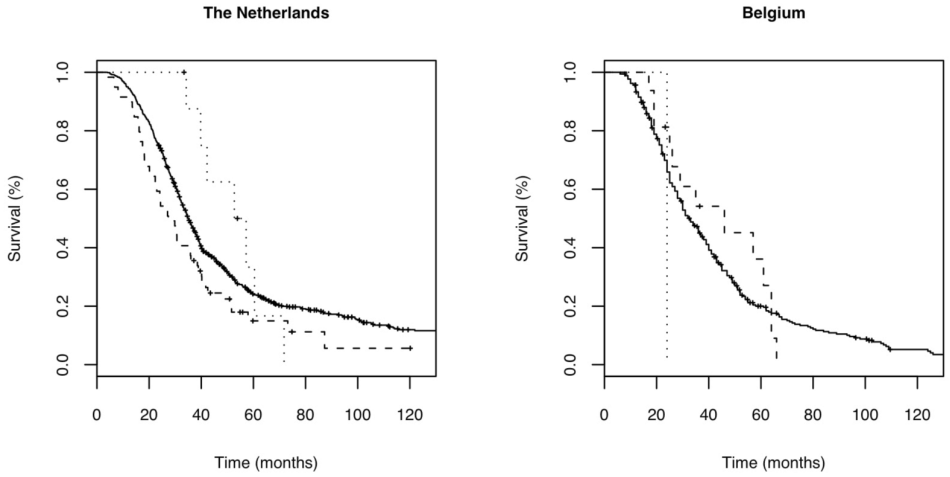


Figure S1: Kaplan-Meier plots showing the survival per study population, for "normal" (uninterrupted line), "short" (dotted line) and "long" (dashed line) NIPA1 polyalanine repeat alleles.

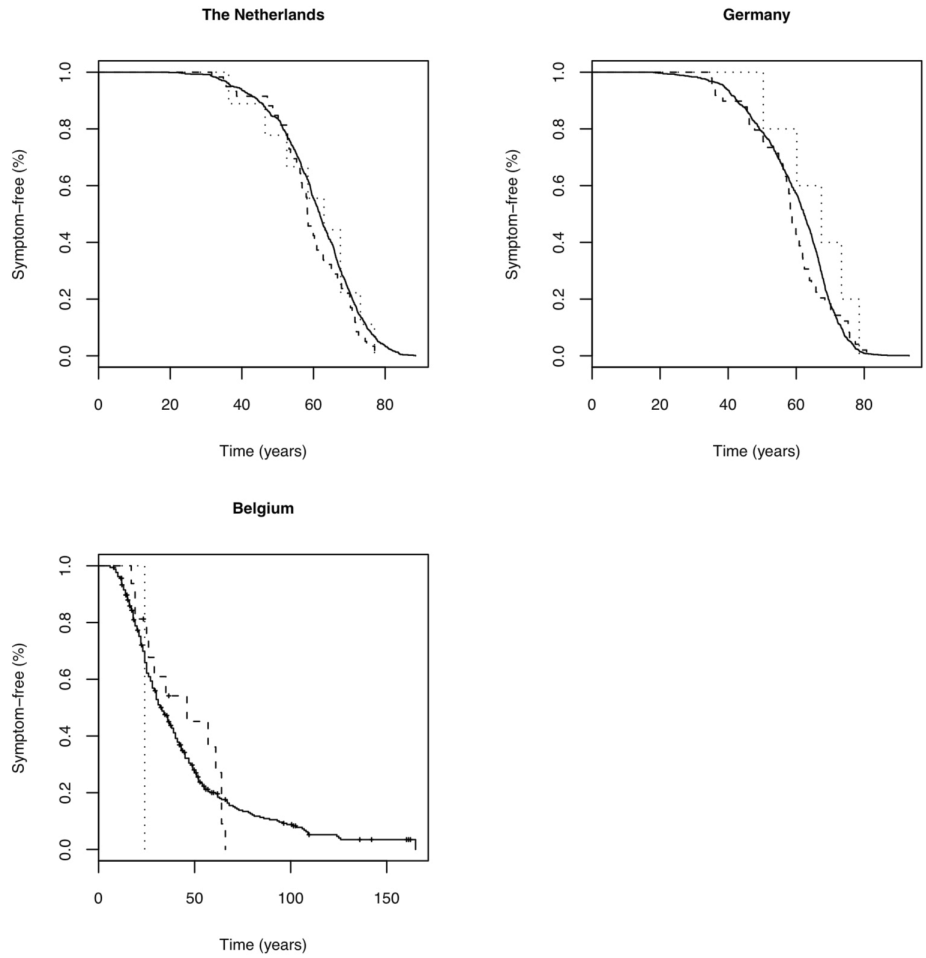


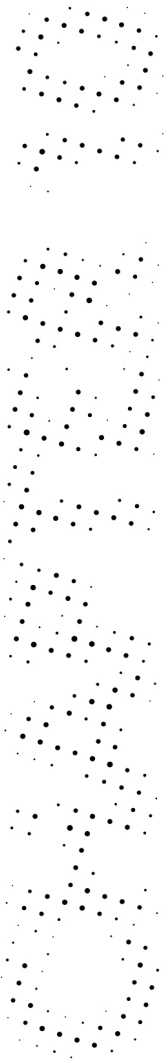
Figure S2: Kaplan-Meier plots showing the age at onset per study population, for "normal" (uninterrupted line), "short" (dotted line) and "long" (dashed line) NIPA1 polyalanine repeat allele

Table S1. *NIPA1* polyalanine repeat allele distribution.

	ALS (%)	Controls (%)
The Netherlands		
(GCG)4	0.2	0.0
(GCG)5	0.1	0.1
(GCG)6	0.3	0.1
(GCG)7	23.1	21.4
(GCG)8	73.2	76.4
(GCG)9	0.7	0.5
(GCG)10	2.5	1.4
(GCG)11	0.0	0.0
	100.0	100.0
Germany		
(GCG)4	0.0	0.0
(GCG)5	0.2	0.1
(GCG)6	0.0	0.1
(GCG)7	21.8	19.9
(GCG)8	75.3	78.6
(GCG)9	0.5	0.2
(GCG)10	2.0	1.2
(GCG)11	0.0	0.0
	100.0	100.0
Belgium		
(GCG)4	0.0	0.0
(GCG)5	0.1	0.1
(GCG)6	0.0	0.4
(GCG)7	22.1	21.5
(GCG)8	75.4	76.2
(GCG)9	0.1	0.6
(GCG)10	2.1	1.2
(GCG)11	0.1	0.0
	100.0	100.0
Combined		
(GCG)4	0.1	0.0
(GCG)5	0.2	0.1
(GCG)6	0.1	0.2
(GCG)7	22.4	21.1
(GCG)8	74.5	76.9
(GCG)9	0.5	0.5
(GCG)10	2.2	1.4
(GCG)11	0.0	0.0
	100.0	100.0

Table S2. Primer sequences.

exon 1	F: 5' - CTCTTCCTGCTCCTCCCCCA
exon 1	R: 5′ - CACCTGCGACCGCCTTCTC
exon 2	F: 5′ - TTCCTGATATACGTAGCTGGT
exon 2	R: 5′ - CCAGACCAGTGATGGATTCT
exon 3	F: 5′ - GTGCTGCGCCCATTTCAGTCA
exon 3	R: 5′ - GTGCCATCTCAACTCACTGCA
exon 4	F: 5′ - AGAAAGGTCAGGTAGTTTGGT
exon 4	R: 5′ - TGGCACTCCTACTGCAAACT
exon 5	F: 5′ - GGGTCTGGTAAATTCAAGCC
exon 5	R: 5′ - CTCCAATGCCTATTCCTCGA
repeat	F: 5′ - [6FAM]GCCCCTCTTCCTGCTCC
repeat	R: 5′ - CGATGCCCTTCTTCTGTAGC



SMN1 GENE DUPLICATIONS ARE ASSOCIATED WITH SPORADIC ALS

Neurology (2012)

Hylke M. Blauw, MD, Christopher Barnes, PhD, Paul W.J. van Vught, PhD, Wouter van Rheenen, MD, Mark Verheul, Edwin Cuppen, PhD, Jan H. Veldink, MD, PhD* Leonard H. van den Berg, MD, PhD*

From the Department of Neurology, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands (H.M.B., P.W.J.V.V., W.V.R., J.H.V., L.H.V.D.B.); Centre for Bioinformatics, Division of Molecular Biosciences and Institute of Mathematical Sciences, Imperial College London, London, UK (C.B.); Hubrecht Institute for Developmental Biology and Stem Cell Research, Royal Netherlands Academy of Sciences (M.V., E.C.) and University Medical Center Utrecht (E.C.), The Netherlands.

* shared last authors.

ABSTRACT

Background: Results of previous relatively small studies that reported on the effect of common copy number variation in the survival motor neuron genes (*SMN1* and *SMN2)* on the risk of sporadic amyotrophic lateral sclerosis (ALS) have been inconsistent. Also, since *SMN* point mutations can cause spinal muscular atrophy (SMA), and the role of *SMN* mutations in ALS has not yet been investigated, we genotyped a large population for sequence and copy number variants.

Methods: We conducted a genetic association study including 847 ALS patients and 984 controls. We used multiplexed ligation-dependent probe amplification (MLPA) assays to determine *SMN1* and *SMN2* copy numbers and examined effects on disease susceptibility and disease course. Furthermore, we sequenced *SMN* genes to determine if *SMN* mutations were more prevalent in ALS patients. A meta-analysis was performed with results from previous studies.

Results: *SMN1* duplications were associated with ALS susceptibility (odds ratio [OR] = 2.07, 95% confidence interval [CI] = 1.34 - 3.20, $p = 0.001$). A meta-analysis with previous data including 3,469 individuals showed a similar effect: OR = 1.85, 95% CI = 1.18 – 2.90, $p = 0.008$). *SMN1* deletions and *SMN2* copy number status were not associated with ALS. *SMN1* or *SMN2* copy number variants had no effect on survival or the age at onset of the disease. We found no enrichment of *SMN* point mutations in ALS patients.

Conclusions: Our data provide firm evidence for a role of common *SMN1* duplications in ALS, and raise new questions regarding the disease mechanisms involved.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease that selectively affects motor neurons in the spinal cord and motor cortex, leading to progressive paralysis and invariably death. The majority of cases have no familial history of the disease and are said to be sporadic. The pathogenesis of sporadic ALS is thought to be an interplay of genetic and environmental risk factors contributing to increased disease susceptibility. The Survival Motor Neuron (*SMN*) gene has been claimed to modulate susceptibility and disease course in ALS.[1] It is present in two copies: *SMN1* and a pseudogene *SMN2,* with about 20% of the biological activity of *SMN1*.[2] Homozygous deletion of *SMN1* causes spinal muscular atrophy (SMA), a congenital motor neuron disease, and higher *SMN2* copy numbers are associated with milder SMA phenotypes.[3] In a minority of cases, SMA is caused by point mutations in *SMN1*, rather than by homozygous deletion.[4]

Because of the phenotypic similarities between SMA and ALS, the role of *SMN* in ALS pathogenesis has been the subject of various studies, several of them reporting significant effects of *SMN1* and *SMN2* copy numbers on disease susceptibility or on disease duration.[1, 5-8] The largest study performed so far showed that abnormal (i.e. 1 and 3) *SMN1* copy numbers were associated with ALS.[8] This investigation did, however, incorporate samples from a prior study, and the results were inconsistent with other reports.[5, 9] Furthermore, the reported effects of *SMN* copy numbers on disease duration have been inconsistent. The role of *SMN* mutations in ALS has never been investigated.

We performed a large-scale association study to determine the effect of *SMN1* and *SMN2* copy numbers on disease susceptibility and on disease course. We genotyped a

genetically homogeneous population, that had not been included in previous reports. In addition, we carried out a comprehensive mutation screen to examine the role of *SMN* coding sequence mutations in ALS

METHODS

ALS patients and healthy volunteers participating in this study were recruited in the outpatient clinic for motor neuron diseases of the Utrecht University Medical Center, or were part of a population-based study on ALS in the Netherlands. This population has been described in detail elsewhere.[10] Patients and controls participating in previous studies on *SMN* were excluded from the copy number analyses. ALS patients had no family history of the disease and all fulfilled the 1994 El Escorial criteria for probable or definite ALS.[11] Genomic DNA of ALS patients and controls was isolated in the same laboratory, using a salting-out procedure. In total we included 847 ALS patients and 984 controls in the copy number analyses and 975 ALS patients and 1044 controls in the mutation screen (Table 1).

Table 1: Patient characteristics

Copy number analyses	ALS	Controls
N (% female)	847 (43)	984 (48)
Mean age (years)	61.9	62.8
Bulbar/ spinal onset (%)	33/ 67	-
Mutation screen	**ALS**	**Controls**
N (% female)	975 (40)	1044 (47)
Mean age (years)	60.4	63.4
Bulbar/ spinal onset (%)	32/68	-

Multiplexed ligation-dependent probe amplification (MLPA) assays were run using standard protocols (www.mlpa.com). We used the SALSA P060 MLPA kit (MRC Holland, The Netherlands), containing 2 probes specifically targeted to *SMN1*, 2 probes targeted to *SMN2*, and control probes targeted to other chromosomal loci for normalization and assay quality control. 50-100 ng of genomic DNA was used in each MLPA assay. Data normalization and analysis were performed with GeneMarker software (SoftGenetics, USA) using standard parameters.

To determine the reproducibility of our MLPA assay, we ran 90 samples twice, in separate reactions and calculated the copy numbers for both replicates of each sample as described below. For the *SMN1* probes, the percentage of agreement was 99% (1 of 90 samples had different copy numbers between the two replicates), and 98% for *SMN2* (2 out of 90 samples showed different copy numbers between replicates).

For mutation screening, we used PCR and sequencing protocols described elsewhere.[12] In short, we designed two nested primer pairs for each amplicon, amplified exonic sequences and intron-exon boundaries and sequenced the amplicons using di-deoxy sequencing. Sequencing was done on ABI 3730 capillary sequencers with Big Dye

Terminator v3.1 chemistry (Applied Biosystems, USA). Sequence data was imported in PolyPhred software[13], and sequences were visually inspected for heterozygous sites. All putative mutations were confirmed with an independent PCR and sequencing reaction. Functional impact of identified mutations was predicted using PolyPhen software (http://genetics.bwh.harvard.edu/pph/). Primer sequences are available upon request. These primers are not specific to *SMN1* or *SMN2*, but amplify sequences from both genes. Identified mutations cannot, therefore, be mapped specifically to one of the two genes. We chose this method because approaches to specifically sequence either *SMN1* or *SMN2* would be extremely laborious, and would only be justified in the case of a suspected association. Two SMA patients with known *SMN1* mutations (in the presence of normal *SMN2* copy numbers) were used as positive controls. The software called both mutations, thus demonstrating that our method reliably detects mutant alleles at least in a 1:3 ratio.

All statistical procedures were carried out in R 2.10.1 statistical environment (http://www.r-project.org). Because the quantitative measurement of copy number data is prone to systematic bias leading to false-positive associations[14] we used two different methods to test *SMN1* and *SMN2* copy number state for association with ALS susceptibility. First, we determined *SMN1* and *SMN2* copy number states for each individual using Gaussian mixture modeling with the CNVtools software package in R.[14] The mean signal of the two probes for each gene was used as the input signal. Gaussian distributions were fitted on the signal intensity distributions and individuals were assigned to copy number states based on the highest a posteriori probability. For *SMN1* a three-component model was used (corresponding to one, two and three copies) and for *SMN2* a five-component model was used (corresponding to zero, one, two, three and four copies). These copy number states were then used in a multivariate logistic regression model including *SMN1* and *SMN2* copy number state and with age at onset and gender as covariates. Secondly, we used a likelihood ratio association test employing CNVtools, using a linear trend model. This method was specifically designed to handle intensity data from quantitative measurements, and allows for differential bias due to possible differences in data quality between cases and controls, causing spurious associations.[14] Cox regression was used to test for effect of *SMN1* and *SMN2* copy number on survival, using age at onset, gender and site of onset as covariates. For the effect on age at onset, we used Cox regression with gender and site of onset as covariates. For the combined analysis of the different studies, we used the random-effects meta-analysis (DerSimonian-Laird) in the rmeta package in R. We used the Woolf test to test for significant heterogeneity between different studies. In order to test for difference in frequency of *SMN* mutations between patients and controls, the Fisher exact test (two-sided) was applied.

Standard protocol approvals, registrations, and patient consents.
The Institutional Review Board and the ethical standards committee at the University Hospital Utrecht approved the studies performed in this project. All patients signed consent forms.

RESULTS

We included 847 ALS patients and 984 controls in the copy number analyses (Table 1). First we tested both *SMN1* and *SMN2* copy number states for association with ALS susceptibility using logistic regression. *SMN1* duplications (i.e. 3 copies) were significantly associated with ALS (odds ratio [OR] = 2.07, 95% confidence interval [CI] = 1.34 - 3.20, $p = 0.001$) (Table 2). There was no effect of *SMN1* deletions (i.e. one *SMN1* copy) on

disease susceptibility (OR = 0.83, 95% CI = 0.42 - 1.65, *p* = 0.60), and we found no effect of *SMN2* copy number states on disease susceptibility. The removal of *SMN2* copy number states from the model that tested *SMN1* and *vice versa*, did not change the results (not shown).

Table 2: *SMN1* and *SMN2* copy number association analysis

	Copy number	ALS n = 847	Controls n = 984	OR (95% CI)	*p*
SMN1	1	16 (1.9%)	23 (2.3%)	0.83 (0.42-1.65)	0.60
	2	771 (91.0%)	926 (94.1%)	.	.
	3	60 (7.1%)	35 (3.6%)	2.07 (1.34-3.20)	1.1×10^{-3}
SMN2	0	62 (7.3%)	78 (7.9%)	0.84 (0.58-1.22)	0.36
	1	329 (38.8%)	372 (37.9%)	1.00 (0.81-1.22)	0.98
	2	416 (49.1%)	486 49.5%)	.	.
	3	39 (4.6%)	46 (4.7%)	0.99 (0.63-1.58)	0.98
	4	1 (0.1%)	2 (0.2%)	0.62 (0.05-7.22)	0.70

OR = odds ratio; CI = confidence interval; *p* = logistic regression p value.

We then used a likelihood ratio test for association to investigate the significance of *SMN* copy numbers on ALS susceptibility. Using this approach, we obtained similar results for both genes: *p* = 0.001 for *SMN1*, *p* = 0.99 for *SMN2*, thus corroborating the results obtained from the logistic regression model.

When combining data from the current study with previously published data, *SMN1* duplications were significantly associated with ALS (OR = 1.85, 95% CI = 1.18 – 2.90, *p* = 0.008) (Figure). There was no significant heterogeneity between different studies (*p* = 0.17, Woolf test). The effect of *SMN1* and *SMN2* deletions was not significant (*SMN1:* OR = 2.23, 95% CI = 0.93 – 5.32, *p* = 0.07; *SMN2:* OR = 1.41, 95% CI = 0.87 – 2.27, *p* = 0.16).

Then the effect of SMN copy number on phenotypic characteristics was tested. Complete clinical data were available for 814 ALS patients. We found no effect of either deletions or duplications of *SMN1* or *SMN2* on disease duration or age at onset (p values > 0.1, not shown).

Sequence data were available for 975 ALS patients and 1044 controls. In our mutation screen we identified 15 heterozygous sequence variants, excluding known single nucleotide polymorphisms (dbSNP build 129) (Table e-1). There was no enrichment of *SMN* mutations in ALS patients: *p* = 0.30, two-tailed Fisher exact test. Two variants, G26D and P198L, will result in amino acid changes. A G26D mutation was identified in one ALS patient; this variant is predicted to be "possibly damaging" by PolyPhen. A P198L mutation was found in a healthy control, and is predicted to be "probably damaging". Both individuals had "normal" (i.e. two) copy number of both *SMN* genes.

Study	ALS	Controls	OR (+/– 95% CI)	*p*
Corcia et al. 2002	7.8% (13/167)	1.8% (3/167)	4.61 (1.29 – 16.51)	
Veldink et al. 2005	10.7% (26/242)	10.3% (18/175)	1.05 (0.56 – 1.98)	
Corcia et al. 2006	5.1% (22/433)	2.6% (12/454)	1.97 (0.96 – 4.03)	
Current study	7.1% (60/847)	3.6% (35/984)	2.07 (1.35 – 3.17)	
Summary	7.2% (121/1689)	3.8% (68/1780)	1.85 (1.18 – 2.90)	0.008

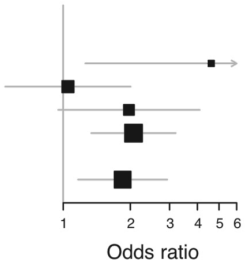


Figure: *combined analysis of SMN1 duplications*

Table shows per study the frequency, absolute numbers and risk estimate of SMN1 duplications. OR = odds ratio; CI = confidence interval; p = logistic regression p value.

Table e-1: discovered (non-polymorphic) sequence variants

Sequence change	Residue	ALS	Controls	Functional impact
5'- CGCGGCACAG (G > A) CCAGGTGAGG -3 '	G26D	1	0	possibly damaging
5'- AGTGCAGTCT (C > T) CCTATTAGCG -3'	-	0	1	noncoding
5'- GGTCAGAAGA (C > T) GGTTGCATTT -3'	D105D	0	1	silent
5'- TCTGTCCGAT (C > T) TACTTTCCCC -3	L141L	2	6	silent
5'- CCTCCACCAC (C > T) CCCCATGCCA -3'	P198L	0	1	probably damaging
5'- CTCCACCACC (C > T) CCCATGCCAG -3'	P198P	1	0	silent
5'- TAGTTAATGT (C > T) GGGACATTTA -3'	-	1	1	noncoding

DISCUSSION

We found a significant effect of *SMN1* duplications on ALS susceptibility, which can be considered a major risk factor for sporadic ALS. The effect size of *SMN1* duplications, obtained from the combined analysis of almost 3,500 individuals, is one of the highest, compared to other established risk factors for ALS. The large sample size and the robust techniques used for data acquisition and analysis, provide confidence in the solidity of the data.

Our results are in line with previous reports that showed that abnormal *SMN1* copy numbers are associated with an increased susceptibility to ALS. In fact, all three studies measuring *SMN1* duplications in the context of ALS reported higher frequency in ALS patients, compared to controls.[6-8] This is reflected by a significant result in the combined analysis, without evidence for heterogeneity between the different studies. With these data, we conclude there is now firm evidence for an association between *SMN1* duplications and ALS susceptibility.

In the current study we find no evidence for an effect of *SMN1* deletions on disease susceptibility. Additionally, the results of the combined analysis weaken previous reported associations of *SMN1* deletions and ALS. The same holds true for the effect of *SMN2* copy number variants, which was not confirmed in our study. One possible explanation for the non-replication of previous reported results is the fact that studies

using quantitative PCR-derived data, such as in copy number studies, are inherently sensitive to different sources of bias.[14] For example, DNA isolation and handling can introduce differential bias between cases and controls, leading to spurious associations.[15] For this reason we used MLPA, a technique that allows simultaneous quantification of multiple probes with one primer pair.[16] This reduces the chances of spurious results due to different PCR reaction properties of target and normalisation primers, as might be obtained with standard quantitative PCR assays. Furthermore, we used a recently developed statistical framework, providing a robust means of association testing of copy number data. The fact that we obtained similar results using different association tests adds to the validity of our data.

In addition to the copy number analyses, we undertook a mutation screen to examine the role of SMN point mutations in ALS pathogenesis, prompted by the fact that a minority of SMA cases are caused by subtle point mutations rather than by gross deletions of *SMN1*. The role of *SMN* mutations in ALS has not been studied before. We find no evidence for a role of point mutations in ALS pathogenesis. Although our approach does not discriminate between mutations in *SMN1* and *SMN2*, and therefore a differential clustering of *SMN1* mutations in ALS patients and *SMN2* mutations in controls cannot be excluded, the lack of enrichment of mutations in ALS patients does make it very unlikely that these mutations contribute significantly to ALS pathogenesis.

The mechanism of *SMN1* duplications on disease susceptibility remains elusive. Given the initial hypothesis that low SMN protein levels increase risk for ALS, in analogy to SMA, this is counterintuitive. An explanation would be that *SMN1* duplications actually produce lower amounts of SMN protein. 5-15% of copy number variants are negatively correlated with gene expression[17] but it is not known if such a relation exists for *SMN*. If a lower amount of SMN protein mediates ALS risk, one would also expect an overrepresentation of *SMN1* deletions in ALS. An alternative and intriguing explanation would be that *SMN* duplications produce higher SMN protein levels that are toxic to motor neurons, but to our knowledge there is as yet no experimental evidence to support this theory yet. Another explanation is that *SMN1* duplications confer a risk factor independent of SMN protein, e.g. because of duplication of other local genomic regions, which we cannot exclude.

In conclusion, our data provide firm evidence that *SMN1* copy number variants are involved in ALS pathogenesis. Further research is needed to explain the increased risk of *SMN1* duplications, rather than deletions. Given the large effect size, as compared to other established risk factors for sporadic ALS, *SMN* duplications are an important risk factor for ALS and further functional studies would be highly justified.

ACKNOWLEDGEMENTS

We would like to thank all patients and healthy volunteers participating in this study. This project was supported by the Prinses 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.

REFERENCES

1. Corcia P, Camu W, Praline J, Gordon P, Vourch P, Andres C. The importance of the SMN genes in the genetics of sporadic ALS. MALS 2009:1-5.
2. Sumner CJ, Huynh TN, Markowitz JA, et al. Valproic acid increases SMN levels in spinal muscular atrophy

patient cells. Ann Neurol 2003;54:647-654.
3. Gavrilov DK, Shi X, Das K, Gilliam TC, Wang CH. Differential SMN2 expression associated with SMA severity. Nat Genet 1998;20:230-231.
4. Rochette CF, Surh LC, Ray PN, et al. Molecular diagnosis of non-deletion SMA patients using quantitative PCR of SMN exon 7. Neurogenetics 1997;1:141-147.
5. Veldink JH, van den Berg LH, Cobben JM, et al. Homozygous deletion of the survival motor neuron 2 gene is a prognostic factor in sporadic ALS. Neurology 2001;56:749-752.
6. Corcia P, Mayeux-Portas V, Khoris J, et al. Abnormal SMN1 gene copy number is a susceptibility factor for amyotrophic lateral sclerosis. Ann Neurol. 2002;51:243-246.
7. Veldink JH, Kalmijn S, Van der Hout AH, et al. SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. Neurology 2005;65:820-825.
8. Corcia P, Camu W, Halimi J-M, et al. SMN1 gene, but not SMN2, is a risk factor for sporadic ALS. Neurology 2006;67:1147-1150.
9. Veldink JH, Kalmijn S, Van der Hout AH, et al. SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. Neurology 2005;65:820-825.
10. van Es MA, van Vught PW, Blauw HM, et al. Genetic variation in DPP6 is associated with susceptibility to amyotrophic lateral sclerosis. Nat Genet 2008;40:29-31.
11. Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors. J Neurol Sci 1994;124 Suppl:96-107.
12. van Boxtel R, Toonen PW, Verheul M, et al. Improved generation of rat gene knockouts by target-selected mutagenesis in mismatch repair-deficient animals. BMC Genomics 2008;9:460.
13. Nickerson DA, Tobe VO, Taylor SL. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res 1997;25:2745-2751.
14. Barnes C, Plagnol V, Fitzgerald T, et al. A robust statistical method for case-control association testing with copy number variation. Nat Genet 2008:8.
15. Craddock N, Hurles ME, Cardin N, et al. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature;464:713-720.
16. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002;30:e57.
17. Stranger BE, Forrest MS, Dunning M, et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 2007;315:848-853.

GENERAL DISCUSSION

In this thesis I describe the results of studies aiming to identify genetic variants that predispose to ALS or modify its disease course. We have used different approaches in order to achieve this: genome-wide SNP association studies (GWAS), CNV association studies exploring both common and rare CNVs, and large-scale candidate gene association studies in which we characterised genetic variation within suspected ALS genes. With these studies we have identified novel genes, which had not been implicated in ALS before, including *UNC13A* and *NIPA1*. We have identified a pathogenic repeat expansion in *NIPA1*, adding further proof that this type of genetic variation is an important theme in ALS. We have shown that large deleterious CNVs with high effect sizes do not play an important role in ALS, in contrast to other traits (1-4). Additionally we have shown that, contrasting previous reports, duplications of *SMN1*, but not *SMN1* deletions or *SMN2* variants, are associated with ALS (5, 6). Together, these studies provide a basis for mechanistic studies to elucidate the pathogenic mechanisms leading to motor neuron degeneration and as a starting point for further hypothesis-generating studies.

Genome-wide association studies (GWASs)

In the last five years there has been a wave of GWASs reporting novel loci that associate to complex traits in health and disease, most of which had not been previously known to play a role in these traits. Currently the online catalogue of published GWASs (7) contains 1350 publications (http://www.genome.gov/gwastudies, accessed 15 august 2012). Although the GWASs have led to many

new discoveries, they were received with considerable scepticism, one of the principal arguments being that the identified loci confer only small effects and thus have minor biological relevance (8). It is certainly true that most GWAS hits confer modest effect sizes: 87% of reported SNPs have ORs of < 2 and only 1% have odds ratios of >10 (9). However, the effect size of a certain locus does not determine its biological importance. For example, an intronic SNP in *HMCGR* (encoding HMG-CoA reductase) confers only a very small effect on plasma LDL cholesterol levels, yet statins, which target this protein, are widely used cholesterol-lowering drugs (10). Another example (from this thesis) is the ALS-associated 9p21.2 locus that confers only a slightly elevated risk (OR=1.15), but this locus harbours the pathogenic repeat expansion in *C9ORF72*, now considered one of the major causes of ALS. So any association, irrespective of the effect size, may be of importance because it may point to important clues about biological pathways contributing to a certain trait. The insight that most common variants will have modest effects, together with lessons from numerous false positive associations, have shaped GWAS study design over the years, which has led to ever larger experimental sample sizes and more rigorous and sophisticated statistical analysis.

The three GWASs described in this thesis mirror this evolution in GWAS design. Our initial GWAS included a total of 1337 ALS patients and 1356 controls and employed a three-stage design. This GWAS identified a SNP that maps to *ITPR2*, a gene involved in glutamatergic neurotransmission. This SNP was significantly associated after Bonferroni correction for multiple testing (corrected p = 0.01). The SNP was significantly associated in each included population, except for the Belgian population and the pooled results showed a p value of $3.3x10^{-6}$. To current standards, however, this result would not be considered significant, using the now generally accepted genome-wide significance level of $p < 5x10^{-8}$. In more recently published studies the association has not been replicated (11, 12). So in spite of the promising results from this study that used a valid design with strict quality control measures and thorough statistical analyses, there is insufficient evidence to consider *ITPR2* a plausible ALS susceptibility gene.

It has been shown that more statistical power can be achieved by using an alternative two-stage study design, including a genome-wide and a replication phase, using a Bonferroni correction for the total number of tested SNPs in the genome-wide phase on the pooled p values from the combined stages (13). Therefore, using this approach, we combined our data with a GWAS from the USA and tested a selected number of SNPs in an independent replication phase. In the pooled analysis the top ranking SNP was significantly associated with $p=5.4x10^{-8}$, (Bonferroni-corrected p= 0.017). The associated SNP showed roughly similar allele frequencies in Dutch, Belgian, Swedish and US populations, and was nominally significant in all but the Belgian populations. The SNP maps to a linkage disequilibrium block with the *DPP6* gene (dipeptidyl peptidase 6), which is mainly expressed in the nervous system (14). Subsequent replication studies in independent populations have yielded confirming (15, 16) but also negative results (11, 12, 17-20). Because of these conflicting results the role of *DPP6* in ALS remains controversial. The contradicting results may indicate a type I error in our initial report. Alternatively, they may be due to genetic heterogeneity in the different study populations (difference in causal genetic variants between different populations). Similarly, if the associated allele actually carries a rare causal variant with high effect size in (low) LD with the associated SNP ("synthetic association"), than the effects will likely be highly population-specific (21). There has been only one report on a resequencing study of *DPP6*, examining only coding regions in a relatively small sample (22). In the case of a

synthetic association, the causal variant may be located far (>1 Mb) from the associated SNP (21). Therefore, deep resequencing of the entire (extended) *DPP6* locus may provide a definitive answer about the role of *DPP6* in ALS.

Of note, from the initial GWASs in ALS it can be concluded that common variants with high effect size do not play an important role in ALS, in contrast to for example in Alzheimer's disease and Creutzfeldt-Jakob disease, where common alleles with high odds ratios have been identified (23, 24). With this knowledge, and based on GWASs in other complex traits where the associated loci confer only modest risk (OR <1.3), we hypothesised that more low-penetrance common variants could be identified when more statistical power was available. We therefore did a large two-staged GWAS that was adequately powered to detect common variants with OR 1.2-1.3 (Chapter 4). In this study, we detected two loci that exceeded the genome-wide significance level of $5x10^{-8}$: one locus on chromosome 19, with *UNC13A* being the only gene in the small area of LD around this SNP, and one locus at chromosome 9p21.2. Both were common variants (as to be expected GWASs are designed to capture common, but not rare variants), with each allele carrying only minimally increased risk: OR = 1.20 and 1.15 for *UNC13A* and 9p21.2 loci, respectively. Both loci showed significant association in the replication phase (Bonferroni corrected $p<0.05$) and exceeded the genome-wide significance level of $5x10^{-8}$.

The *UNC13A* locus has not been implicated in ALS before. Since our initial report, there have been two other reports that replicated the association and two studies that did not (11, 25-27). In addition, the identified SNP was associated with a shorter survival in two reports that included Dutch and Italian patients (25, 26). Although *UNC13A* is the closest (and only) mapping gene in the associated LD block and is of biological relevance (12), the causal variant has not been identified yet and the biological mechanism that links the associated locus to neurodegeneration remains to be determined.

The associated locus at chromosome 9p21.1 is included in a linkage region that was identified in pedigrees with ALS and frontotemporal dementia (FTD) (28). The results from our GWAS allowed for a considerable finemapping of this linkage region spanning several megabases. In 2011 two reports showed that the causal variant in these families is an intronic hexanucleotide repeat mutation (HREM) in *C9ORF72*, which maps to this 9p21.1 locus (29, 30). This HREM is now known to explain 23-84% of familial ALS and 80-100% of familial ALS with FTD, especially in northern European countries (31). The HREM is also detected in 6-8% of sporadic ALS patients, while it is virtually absent (0-0.4%) in controls (32). It has been shown that almost all HREM occur on a common risk haplotype that is tagged by rs3849942, one of the most significantly associated SNPs in our GWAS (30, 31). So the identified association of the 9p21.2 locus in our GWAS is likely to be explained by the HREM in *C9ORF72,* although it cannot be completely excluded that an additional effect conferred by the associated SNP is present. If the association is indeed mediated by the HREM, than this could be regarded as a good example of a synthetic association where a common variant is detected because of a rare variant residing on the associated risk haplotype. While most identified GWAS association signals are expected to be explained by a common variant that is tagged by the associated SNP, a (modest) proportion of identified GWAS hits may represent such synthetic associations (21, 33, 34). It is illustrative that the causal variant confers a high risk (OR = 25.7), while the SNP tagging the risk allele confers only a modest increase in risk (OR = 1.15), thus explaining why this variant was not detected in our earlier, relatively small GWASs.

CNV studies

Because deletions and insertions of genomic sequence (copy-number variants, CNVs) can have a major impact on (disease) phenotypes, we sought to systematically assess their role as possible mediators of ALS liability. Apart from anecdotal reports, the role of CNVs in ALS had not been studied previously. Because SNP arrays generate intensity data for each probe (in addition to genotype information), it was possible to use the data from our GWAS for CNV detection. Because there was no automated CNV detection software we developed an algorithm to analyse the SNP intensity profiles of the samples that had been genotyped for the earlier GWAS. We examined CNV characteristics of patients and controls and did an association analysis to identify potential risk loci (Chapter 5). In a second CNV study (Chapter 6), we combined our data with an Irish dataset, and sought to identify possible genes of interest by examining ALS-specific CNVs that were observed in both populations.

An important observation from these two studies is that we did not identify an increased CNV burden in ALS patients, compared to controls, as suggested by others (35). Furthermore, we did not identify associated loci where large-scale, deleterious CNVs are more frequent in ALS patients than in controls. In agreement with other studies (36) most CNVs were rare and many were private (i.e. found in only one subject in the study population), precluding meaningful association analyses. Because we hypothesised that CNVs might differ at group level, we examined the genic content for CNVs in patients and controls. In the Dutch population we found an enrichment of ALS-specific deleted genes, indicating that while no individual associated variant could be identified, a mosaic of many different private CNVs might influence ALS liability by their genic content. However, this finding could not be replicated in the Irish population or the larger study described in Chapter 7.

In addition to the relatively small exploratory CNV studies, we sought to perform a large-scale association study in order to identify rare ALS-associated CNVs. Because statistical power is heavily dependent on the allele frequency of the studied variants, a large sample size is needed to detect meaningful associations, which is of particular relevance for CNVs, which are typically rare. In a collaborative effort, we did a two-stage study including more than 19,000 individuals from different countries. In our initial discovery phase we identified two nominally significant CNV loci, which we tested in an independent replication experiment (Chapter 7).

One of the two tested loci in our replication study was the *DPP6* locus, the strongest associated gene in our second (SNP-based) GWAS. The finding that rare CNVs at this locus associate to ALS was therefore an intriguing finding. It must be noted that the discovery phase in our CNV study included the populations that had been used in the GWAS that produced the SNP association, raising the question whether these results represent independent findings. However, because of the distance between the CNVs and the associated SNP, and the absence of significant LD, it seems plausible that the two observations are independent and therefore add support to *DPP6* as a susceptibility gene for ALS. The other locus that was tested in the replication phase was the 15q11.2 region, containing *NIPA1*, among others. *NIPA1* is of biological relevance because mutations cause hereditary spastic paraplegia (HSP) type 6, characterised by slowly progressive, mainly upper-motor neuron degeneration (37).

For both loci the results of the replication study were not significant after correction for multiple testing (*DPP6* locus: nominal p = 0.097, 15q11.2: p = 0.041, Fisher's exact test). However, CNVs were more frequent in patients than in controls in both loci (*DPP6*: 0.39% vs 0.20%, *NIPA1*: 0.27% vs 0.08%), and the pooling of the discovery and replication phases led to lower p values, suggesting that the replication phase was underpowered.

Therefore, we conclude that although there is not sufficient evidence to claim associations of either locus with ALS, they are candidates for follow-up studies (see below).

In these studies we unavoidably biased against smaller, more common CNVs (copy-number polymorphisms, CNPs). This is inherent to the used methodology: polymorphic CNV regions typically do not pass SNP quality control filters and therefore usually have been excluded from the first generation SNP genotyping platforms. Additionally, the typically employed CNV calling procedures in CNV studies bias against small CNVs because they need several consecutive markers with deviation in probe intensity to call a CNV. Therefore, CNPs as a group have not been systematically assessed for most traits and have not been included in our previous CNV studies (38-40).

To further explore the role of CNPs in ALS we used a novel approach: we tested a set of previously validated CNPs in an association study using the CNV370 array, which includes probes that are specifically targeted at CNP regions. We tested 253 CNPs from which we selected the most associated CNPs and tested these in an independent population. In this replication experiment we tested 5 CNPs (4 associated with ALS susceptibility, 1 associated with a shorter survival), but none reached statistical significance. Therefore, we can conclude that the tested set of CNPs do not contain ALS-associated CNPs. This is in agreement with a large CNP association study, which found only a modest contribution of CNPs to the eight traits that were studied (41).

The systematic genome-wide investigation of CNPs in association studies remains a challenge. While large, rare CNVs can now be ascertained straightforwardly in large case-control cohorts using SNP arrays, the genotyping of CNPs is notoriously difficult (42). To our knowledge, only one large-scale genome-wide case-control study employing CNP genotyping, has been published, probably reflecting this difficulty (41). Different kinds of technical issues can produce spurious results and false associations (41-43). In addition, the content of older and even current SNP genotyping platforms may not be adequate to capture most CNPs (41, 44, 45). Especially challenging are the multiallelic, complex CNPs, which are enriched in segmental duplications: duplicated genomic regions of highly similar sequence content. These regions harbour several disease-associated CNPs (46-48) but due to a paucity of tagSNPs in these regions are not captured well in (SNP) GWASs (44). Even in a dedicated CNP association study, the genotyping of these CNPs can be challenging (41, 44). The *SMN* region, characterised by segmental duplications, serves as an example to illustrate this (Chapter 10): the older SNP arrays (used in our initial GWAs) do not contain probes that map to this region, so cannot be explored indirectly via tagSNPs. The newer CNV370 array includes several probes in this region, but these probes do not differentiate between *SMN1* and *SMN2*, and thus are not informative. To accurately genotype these genes we had to use MLPA assays, specifically designed to assay this particular locus. Extending from this observation, for a comprehensive genome-wide survey of all CNVs (CNPs) a dedicated platform and study design may be needed. However, it is questionable whether the costs for this are justified, and perhaps it is better to await the arrival of whole-genome sequencing case control studies to provide the final answer.

NIPA1

Because of our finding of a possible role of *NIPA1* deletions in ALS, we investigated the role of sequence variants in *NIPA1* to further characterise its role in ALS. We sequenced the coding regions in a large association study including ~2,300 patients and ~2,800 controls (Chapter 9). We found two damaging mutations in ALS patients; no damaging mutations were found in controls. Although the low frequency of these variants precludes meaningful statistical analyses to prove or disprove any association with ALS, *NIPA1* point

mutation could be a rare cause of ALS.

In addition to the sequencing experiments, we genotyped a polyalanine repeat in the first exon of *NIPA1*. We found that long repeat alleles were associated with ALS susceptibility (OR = 1.71, $p = 1.6x10^{-4}$), duration of survival (HR = 1.59, $p = 4.2x10^{-4}$) and age at onset of disease (HR = 1.39, $p = 4.6x10^{-3}$). In addition to the suggestive evidence from our CNV study implicating *NIPA1* deletions (Chapter 7), these findings support the role of *NIPA1* in ALS susceptibility and as a modifier of disease course.

While *NIPA1* mutations are known to cause hereditary spastic paraplegia (HSP) type 6, our reports were the first to implicate *NIPA1* in ALS pathology. The observation that variants in the same gene, depending on the genetic and environmental context, lead to different phenotypes ("pleiotropy") is also seen with other ALS-associated genes. For example, *ATXN2* mutations are implicated in ALS and in spinocerebellar ataxia, *C9ORF72* mutations can lead to ALS and frontotemporal dementia, and *ANG* variants are implicated in ALS and Parkinson's disease (31, 49, 50). This suggests that these clinical entities have pathogenic pathways in common, which could have implications for further genetic studies. Intriguingly, a recent neuropathological report describes marked TDP-43 staining, the pathological hallmark of ALS, in a patient with HSP type 6 caused by a *NIPA1* mutation (51). This suggests that *NIPA1* may be involved in important pathogenic pathways that are shared by HSP and ALS.

With *ATXN2* and *C9ORF72*, *NIPA1* is the third gene that has been proposed to associate to ALS repeat-length variation. Of note, these three genes have been identified in candidate gene studies and only one (*C9ORF72*) produced a GWAS signal, which raises the question how many other ALS-associated repeat expansions have yet to be identified. Only one study studied seven disease-related polyglutamine expansions in a modest sample, but found no associations with ALS (52). Possibly, a comprehensive survey of all potentially pathogenic repeats might reveal other ALS-associated repeat expansions.

CONCLUSION AND FUTURE PERSPECTIVES

Like in other complex traits, the allelic architecture that underlies (sporadic) ALS susceptibility appears to be a spectrum of rare variants conferring a high risk and common variants conferring low risks (Figure). There are probably no common, high-risk variants that predispose to ALS, because they would have been detected in the various GWASs: the GWAS described in Chapter 4 had 100% power to detect variants of OR > 2, and the probes on the used SNP array capture >90% of common alleles with r^2 of >0.5 (www.Illumina.com, technical bulletin). In contrast to diseases such as schizophrenia, autism and idiopathic epilepsy, rare deleterious CNVs do not play a major role in ALS.

With the increasing availability of whole-genome sequencing, it seems reasonable to assume that most (if not all) familial ALS genes will be identified within the next years. Recently, several causative genes in Mendelian diseases, including familial ALS, have been uncovered this way (53). However, the unravelling of the genetic basis of sporadic ALS has proven to be challenging. So far, with the exception of a number of familial ALS genes, only a few loci (with modest effects) have been discovered that convincingly contribute to sporadic ALS (12) and these loci do not explain the heritability of 0.61 as calculated from twin data (54). So, as is currently the case for most complex traits, there will be many risk loci that are yet to be identified. The issue of where this "missing heritability" is to be found has been heavily debated in the literature and is of importance for the design of future studies (55). The various hypotheses regarding this issue include

(among others) the role of multiple common variants with effects that are too small to be detected in current GWASs (56, 57) and the role of multiple rare variants with large effects that are not captured by GWASs (58). Both will likely contribute to ALS liability and thus represent subjects for future study, each requiring different approaches.

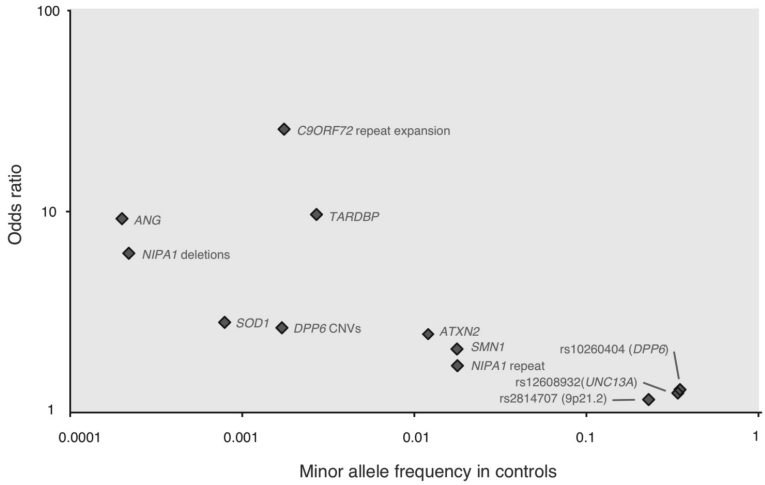


Figure: *genetic variants that have been associated with sporadic ALS.*
For each gene/ variant the frequency in controls and the effect size (odds ratio) is shown. Please note that the figure serves to illustrate the spectrum of allele frequencies and associated risk for each variant, rather than as an exhaustive review of ALS-associated genes. References: (12, 31, 40, 48-50, 65-67)

Larger, smarter GWASs

There is evidence that a substantial proportion of missing heritability is explained by common variants that are either in low LD with the trait-associated causal variants or have effects that are too small to be detected in moderately powered GWASs (57). It is becoming somewhat of a cliché to propose that further increase of sample size of future GWASs will lead to the identification of more novel associations, but there is no doubt that this is true, also for ALS. As demonstrated by GWASs in other traits, the increase in sample size of study populations was almost linearly related to the number of newly identified associations (8). The largest GWAS in ALS that has been published included about 4,800 patients and 15,000 controls (12), which is impressive compared to the initial studies, but modest compared to studies done in for example ulcerative colitis (>16,000 cases, >32,000 controls), Crohn's disease (>22,000 cases, 29,000 controls) and type 2 diabetes (35,000 cases, 115,000 controls), producing an equally impressive number of identified risk loci (59-61). So, based on theoretical and empirical data, it can be assumed that many more associated common variants can be identified with larger study populations.

Alternatively, perhaps much can be gained by classifying subjects based on endophenotypes, such as gene expression profiles, extreme traits (e.g. extreme long or short survival), specific neuropathological features or other biomarkers. It has been shown that ALS is a genetically heterogeneous disease (31, 62), which will reduce power to detect associations. Hence, classification that will lead to more homogeneous

populations may lead to the identification of novel loci, as illustrated by a GWAS in frontotemporal dementia (63).

Rare variants, repeats, complex CNVs

Rare genetic variants likely play an important role in most complex traits, as demonstrated by theoretical studies and several empirical examples, including ALS studies (30, 50). Because GWASs are designed to capture common alleles, most rare alleles will go undetected with this approach. The genome-wide assessment of rare sequence variants is currently still problematic due to technical limitations. However, in the (near) future, case-control studies employing whole-genome sequencing technology will likely become feasible. However, these will require perhaps even larger sample sizes than for SNP GWASs (64). Whole genome sequencing may also be the technique of choice to interrogate more complex genomic regions containing repetitive sequences and complex CNVs, which are the least likely to be detected in LD-based association studies and are difficult to characterise using array-based techniques (41).

Finally, explaining heritability is not a goal in itself. The main goal of genetic studies should be to identify as much variants as needed to understand the biological pathways that lead to motor neuron degeneration and how to manipulate these by therapeutic strategies. The ultimate goal will be to translate genetic discoveries into clinical applications and to discover an effective treatment for patients suffering from this devastating disease.

REFERENCES

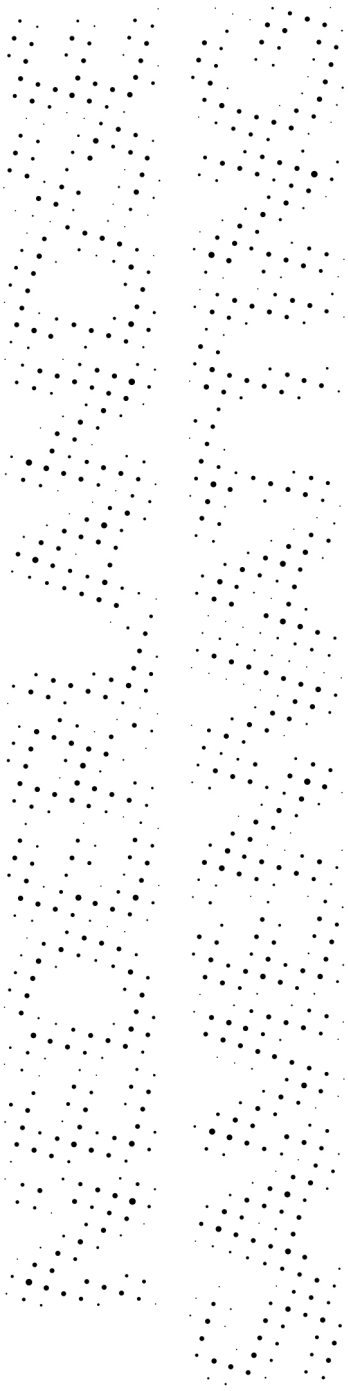
1. Stone, J.L., O'donovan, M.C., Gurling, H., Kirov, G.K., Blackwood, D.H.R., Corvin, A., Craddock, N.J., Gill, M., Hultman, C.M., Lichtenstein, P. *et al.* (2008) Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, 455, 237-241.
2. De Kovel, C.G.F., Trucks, H., Helbig, I., Mefford, H.C., Baker, C., Leu, C., Kluck, C., Muhle, H., Von Spiczak, S., Ostertag, P. *et al.* (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain*, 133, 23-32.
3. Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A.R., Green, T. *et al.* (2008) Association between microdeletion and microduplication at 16p11.2 and autism. *N Engl J Med*, 358, 667-675.
4. Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P.H., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E. *et al.* (2008) Large recurrent microdeletions associated with schizophrenia. *Nature*, 455, 232-236.
5. Veldink, J.H., Kalmijn, S., Van der Hout, A.H., Lemmink, H.H., Groeneveld, G.J., Lummen, C., Scheffer, H., Wokke, J.H.J. and Van den Berg, L.H. (2005) SMN genotypes producing less SMN protein increase susceptibility to and severity of sporadic ALS. *Neurology*, 65, 820-825.
6. Corcia, P., Camu, W., Halimi, J.-M., Vourc'h, P., Antar, C., Vedrine, S., Giraudeau, B., de Toffol, B., Andres, C.R. and Group, F.A.S. (2006) SMN1 gene, but not SMN2, is a risk factor for sporadic ALS. *Neurology*, 67, 1147-1150.
7. Hindorff, L.A., Sethupathy, P., Junkins, H.A., Ramos, E.M., Mehta, J.P., Collins, F.S. and Manolio, T.A. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*, 106, 9362-9367.
8. Visscher, Peter M., Brown, Matthew A., Mccarthy, Mark I. and Yang, J. (2012) Five Years of GWAS Discovery. *The American Journal of Human Genetics*, 90, 7-24.
9. Cirulli, E.T. and Goldstein, D.B. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet*, 11, 415-425.

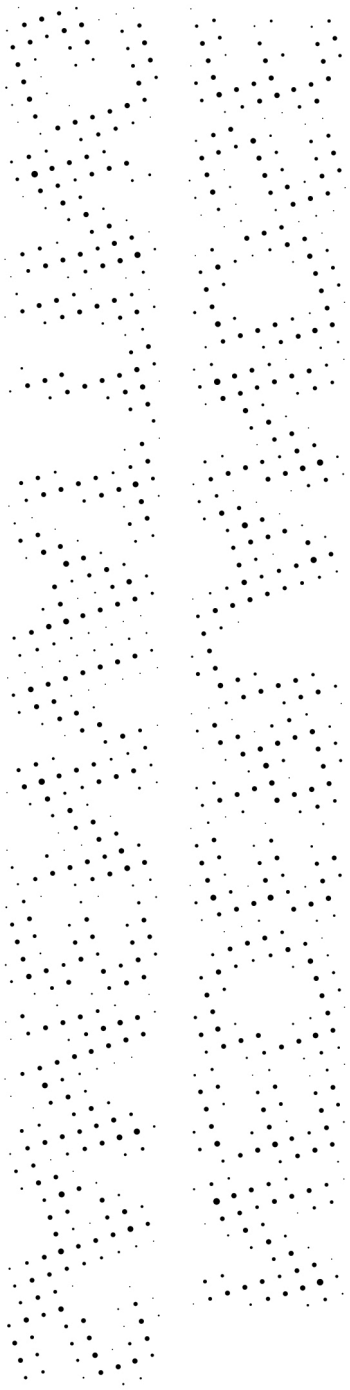
10. Kathiresan, S., Willer, C.J., Peloso, G.M., Demissie, S., Musunuru, K., Schadt, E.E., Kaplan, L., Bennett, D., Li, Y., Tanaka, T. *et al*. (2009) Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat Genet*, 41, 56-65.
11. Shatunov, A., Mok, K., Newhouse, S., Weale, M.E., Smith, B., Vance, C., Johnson, L., Veldink, J.H., van Es, M.A., van den Berg, L.H. *et al*. (2010) Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol*, 9, 986-994.
12. Van Es, M.A., Veldink, J.H., Saris, C.G.J., Blauw, H.M., Van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B. *et al*. (2009) Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet*, 41, 1083-1087.
13. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet*, 38, 209-213.
14. Clark, B.D., Kwon, E., Maffie, J., Jeong, H.Y., Nadal, M., Strop, P. and Rudy, B. (2008) DPP6 Localization in Brain Supports Function as a Kv4 Channel Associated Protein. *Front Mol Neurosci*, 1, 8.
15. Bo, R.D., Ghezzi, S., Corti, S., Santoro, D., Prelle, A., Mancuso, M., Siciliano, G., Briani, C., Murri, L., Bresolin, N. *et al*. (2008) DPP6 gene variability confers increased risk of developing sporadic amyotrophic lateral sclerosis in Italian patients. *Journal of Neurology, Neurosurgery & Psychiatry*, 79, 1085-1085.
16. Cronin, S., Berger, S., Ding, J., Schymick, J.C., Washecka, N., Hernandez, D.G., Greenway, M.J., Bradley, D.G., Traynor, B.J. and Hardiman, O. (2008) A genome-wide association study of sporadic ALS in a homogenous Irish population. *Human Molecular Genetics*, 17, 768-774.
17. Chiò, A., Schymick, J.C., Restagno, G., Scholz, S.W., Lombardo, F., Lai, S.-L., Mora, G., Fung, H.-C., Britton, A., Arepalli, S. *et al*. (2009) A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum Mol Genet*, 18, 1524-1532.
18. Cronin, S., Tomik, B., Bradley, D.G., Slowik, A. and Hardiman, O. (2009) Screening for replication of genome-wide SNP associations in sporadic ALS. *Eur J Hum Genet*, 17, 213-218.
19. Fogh, I., D'Alfonso, S., Gellera, C., Ratti, A., Cereda, C., Penco, S., Corrado, L., Soraru, G., Castellotti, B., Tiloca, C. *et al*. (2009) No association of DPP6 with amyotrophic lateral sclerosis in an Italian population. *Neurobiol Aging*, 32, 966-967.
20. Landers, J.E., Melki, J., Meininger, V., Glass, J.D., van den Berg, L.H., Van Es, M.A., Sapp, P.C., Van Vught, P.W.J., McKenna-Yasek, D.M., Blauw, H.M. *et al*. (2009) Reduced expression of the Kinesin-Associated Protein 3 (KIFAP3) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proc Natl Acad Sci USA*, 106, 9004-9009.
21. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. and Goldstein, D.B. (2010) Rare Variants Create Synthetic Genome-Wide Associations. *PLoS Biol*, 8, e1000294.
22. Daoud, H., Valdmanis, P.N., Dion, P.A. and Rouleau, G.A. (2009) Analysis of DPP6 and FGGY as candidate genes for amyotrophic lateral sclerosis. *Amyotroph Lateral Scler*, 11, 389-391.
23. Mead, S., Poulter, M., Uphill, J., Beck, J., Whitfield, J., Webb, T.E.F., Campbell, T., Adamson, G., Deriziotis, P., Tabrizi, S.J. *et al*. (2009) Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *Lancet neurology*, 8, 57-66.
24. Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S. and Roses, A.D. (1993) Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proc Natl Acad Sci U S A*, 90, 1977-1981.
25. Chio, A., Mora, G., Restagno, G., Brunetti, M., Ossola, I., Barberis, M., Ferrucci, L., Canosa, A., Manera, U., Moglia, C. *et al*. (2012) UNC13A influences survival in Italian amyotrophic lateral sclerosis patients: a population-based study. *Neurobiol Aging*.
26. Diekstra, F.P., van Vught, P.W., van Rheenen, W., Koppers, M., Pasterkamp, R.J., van Es, M.A., Schelhaas, H.J., de Visser, M., Robberecht, W., Van Damme, P. *et al*. (2011) UNC13A is a modifier of survival in amyotrophic lateral sclerosis. *Neurobiol Aging*, 33, 630 e633-638.
27. Daoud, H., Belzil, V., Desjarlais, A., Camu, W., Dion, P.A. and Rouleau, G.A. (2010) Analysis of the UNC13A gene as a risk factor for sporadic amyotrophic lateral sclerosis. *Arch Neurol*, 67, 516-517.
28. Vance, C., Al-Chalabi, A., Ruddy, D., Smith, B.N., Hu, X., Sreedharan, J., Siddique, T., Schelhaas, H.J., Kusters, B., Troost, D. *et al*. (2006) Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain*, 129, 868-876.

29. Renton, A.E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., Van Swieten, J.C., Myllykangas, L. *et al.* (2011) A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron*, 72, 257-268.
30. Dejesus-Hernandez, M., Mackenzie, Ian R., Boeve, Bradley F., Boxer, Adam L., Baker, M., Rutherford, Nicola J., Nicholson, Alexandra M., Finch, NiCole A., Flynn, H., Adamson, J. *et al.* (2011) Expanded GGGGCC Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron*, 1-13.
31. Smith, B.N., Newhouse, S., Shatunov, A., Vance, C., Topp, S., Johnson, L., Miller, J., Lee, Y., Troakes, C., Scott, K.M. *et al.* (2012) The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur J Hum Genet*, Epub ahead of print.
32. van Rheenen, W., van Blitterswijk, M., Huisman, M.H., Vlam, L., van Doormaal, P.T., Seelen, M., Medic, J., Dooijes, D., de Visser, M., van der Kooi, A.J. *et al.* (2012) Hexanucleotide repeat expansions in C9ORF72 in the spectrum of motor neuron diseases. *Neurology*, 79, 878-882.
33. Wray, N.R., Purcell, S.M. and Visscher, P.M. (2011) Synthetic Associations Created by Rare Variants Do Not Explain Most GWAS Results. *PLoS Biol*, 9, e1000579.
34. Orozco, G., Barrett, J.C. and Zeggini, E. (2010) Synthetic associations in the context of genome-wide association scan signals. *Hum Mol Genet*, 19, R137-R144.
35. Meyer, T., Alber, B., Roemer, K., Martin, T., Kalscheuer, V.M., Göttert, E., Zang, K.D., Ludolph, A.C., Ropers, H.-H. and Prudlo, J. (2003) High rate of constitutional chromosomal rearrangements in apparently sporadic ALS. *Neurology*, 60, 1348-1350.
36. Redon, R., Ishikawa, S., Fitch, K.R., Feuk, L., Perry, G.H., Andrews, T.D., Fiegler, H., Shapero, M.H., Carson, A.R., Chen, W. *et al.* (2006) Global variation in copy number in the human genome. *Nature*, 444, 444-454.
37. Rainier, S., Chai, J.-H., Tokarz, D., Nicholls, R.D. and Fink, J.K. (2003) NIPA1 gene mutations cause autosomal dominant hereditary spastic paraplegia (SPG6). *Am J Hum Genet*, 73, 967-971.
38. Cronin, S., Blauw, H.M., Veldink, J.H., van Es, M.A., Ophoff, R.A., Bradley, D.G., van den Berg, L.H. and Hardiman, O. (2008) Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. *Hum Mol Genet*, 17, 3392-3398.
39. Blauw, H.M., Veldink, J.H., Van Es, M.A., Van Vught, P.W., Saris, C.G.J., van der Zwaag, B., Franke, L., Burbach, J.P.H., Wokke, J.H., Ophoff, R.A. *et al.* (2008) Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. *Lancet neurology*, 7, 319-326.
40. Blauw, H.M., Al-Chalabi, A., Andersen, P.M., van Vught, P.W., Diekstra, F.P., van Es, M.A., Saris, C.G., Groen, E.J., van Rheenen, W., Koppers, M. *et al.* (2010) A large genome scan for rare CNVs in amyotrophic lateral sclerosis. *Hum Mol Genet*, 19, 4091-4099.
41. Craddock, N., Hurles, M.E., Cardin, N., Pearson, R.D., Plagnol, V., Robson, S., Vukcevic, D., Barnes, C., Conrad, D.F., Giannoulatou, E. *et al.* (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature*, 464, 713-720.
42. Barnes, C., Plagnol, V., Fitzgerald, T., Redon, R., Marchini, J., Clayton, D. and Hurles, M.E. (2008) A robust statistical method for case-control association testing with copy number variation. *Nat Genet*, 40, 1245-1252.
43. Aldhous, M.C., Abu Bakar, S., Prescott, N.J., Palla, R., Soo, K., Mansfield, J.C., Mathew, C.G., Satsangi, J. and Armour, J.A.L. (2010) Measurement methods and accuracy in copy number variation: failure to replicate associations of beta-defensin copy number with Crohn's disease. *Hum Mol Genet*, 19, 4930-4938.
44. Campbell, C.D., Sampas, N., Tsalenko, A., Sudmant, P.H., Kidd, J.M., Malig, M., Vu, T.H., Vives, L., Tsang, P., Bruhn, L. *et al.* (2011) Population-genetic properties of differentiated human copy-number polymorphisms. *Am J Hum Genet*, 88, 317-332.
45. McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet*, 40, 1166-1174.
46. Fanciulli, M., Norsworthy, P.J., Petretto, E., Dong, R., Harper, L., Kamesh, L., Heward, J.M., Gough, S.C., de Smith, A., Blakemore, A.I. *et al.* (2007) FCGR3B copy number variation is associated with susceptibility to systemic, but not organ-specific, autoimmunity. *Nat Genet*, 39, 721-723.

47. Hollox, E.J., Huffmeier, U., Zeeuwen, P.L.J.M., Palla, R., Lascorz, J., Rodijk-Olthuis, D., Van De Kerkhof, P.C.M., Traupe, H., De Jongh, G., Heijer, M.D. *et al.* (2008) Psoriasis is associated with increased β-defensin genomic copy number. *Nat Genet*, 40, 23-25.
48. Blauw, H.M., Barnes, C.P., van Vught, P.W., van Rheenen, W., Verheul, M., Cuppen, E., Veldink, J.H. and van den Berg, L.H. (2012) SMN1 gene duplications are associated with sporadic ALS. *Neurology*, 78, 776-780.
49. Elden, A.C., Kim, H.-J., Hart, M.P., Chen-Plotkin, A.S., Johnson, B.S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M.M. *et al.* (2010) Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, 466, 1069-1075.
50. Van Es, M.A., Schelhaas, H.J., Van Vught, P.W.J., Ticozzi, N., Andersen, P.M., Groen, E.J.N., Schulte, C., Blauw, H.M., Koppers, M., Diekstra, F.P. *et al.* (2011) Angiogenin variants in Parkinson disease and amyotrophic lateral sclerosis. *Ann Neurol.*, 70, 964-973.
51. Martinez-Lage, M., Molina-Porcel, L., Falcone, D., McCluskey, L., Lee, V.M.-Y., Van Deerlin, V.M. and Trojanowski, J.Q. (2012) TDP-43 pathology in a case of hereditary spastic paraplegia with a NIPA1/SPG6 mutation. *Acta neuropathologica*.
52. Lee, T., Li, Y.R., Chesi, A., Hart, M.P., Ramos, D., Jethava, N., Hosangadi, D., Epstein, J., Hodges, B., Bonini, N.M. *et al.* (2011) Evaluating the prevalence of polyglutamine repeat expansions in amyotrophic lateral sclerosis. *Neurology*, 76, 2062-2065.
53. Wu, C.-H., Fallini, C., Ticozzi, N., Keagle, P.J., Sapp, P.C., Piotrowska, K., Lowe, P., Koppers, M., Mckenna-Yasek, D., Baron, D.M. *et al.* (2012) Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. *Nature*, 1-7.
54. Al-Chalabi, A., Fang, F., Hanby, M.F., Leigh, P.N., Shaw, C.E., Ye, W. and Rijsdijk, F. (2010) An estimate of amyotrophic lateral sclerosis heritability using twin data. *J Neurol Neurosurg Psychiatry*, 81, 1324-1326.
55. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., Mccarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, 461, 747-753.
56. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'donovan, M.C., Sullivan, P.F., Sklar, P., Purcell (Leader, S.M., Stone, J.L., Sullivan, P.F. *et al.* (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 1-5.
57. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W. *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*, 42, 565-569.
58. Pritchard, J.K. (2001) Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet*, 69, 124-137.
59. Anderson, C.A., Boucher, G., Lees, C.W., Franke, A., D'Amato, M., Taylor, K.D., Lee, J.C., Goyette, P., Imielinski, M., Latiano, A. *et al.* (2011) Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet*, 43, 246-252.
60. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Magi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S. *et al.* (2012) Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat Genet*, Epub ahead of print.
61. Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balschun, T., Lee, J., Roberts, R. *et al.* (2010) Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet*, 42, 1118-1125.
62. van Es, M.A., Dahlberg, C., Birve, A., Veldink, J.H., van den Berg, L.H. and Andersen, P.M. (2009) Large-scale SOD1 mutation screening provides evidence for genetic heterogeneity in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*, 81, 562-566.
63. Van Deerlin, V.M., Sleiman, P.M.A., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.-S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M. *et al.* (2010) Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet*, 42, 234-239.
64. Kiezun, A., Garimella, K., Do, R., Stitziel, N.O., Neale, B.M., Mclaren, P.J., Gupta, N., Sklar, P., Sullivan, P.F., Moran, J.L. *et al.* (2012) Exome sequencing and the genetic basis of complex traits. *Nat Genet*, 44, 623-630.
65. Blauw, H.M., Van Rheenen, W., Koppers, M., Van Damme, P., Waibel, S., Lemmens, R., Van Vught, P.W.J., Meyer,

T., Schulte, C., Gasser, T. *et al.* (2012) NIPA1 polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. *Hum Mol Genet*, 21, 2497-2502.
66. Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande Velde, C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F. *et al.* (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. *Nat Genet*, 40, 572-574.
67. Van Blitterswijk, M., Van Es, M.A., Hennekam, E.A.M., Dooijes, D., Van Rheenen, W., Medic, J., Bourque, P.R., Schelhaas, H.J., Van Der Kooi, A.J., De Visser, M. *et al.* (2012) Evidence for an oligogenic basis of amyotrophic lateral sclerosis. *Hum Mol Genet*, 21, 3776-3784.





Amyotrofische laterale sclerose (ALS) is een neurodegeneratieve ziekte die in Nederland jaarlijks enkele honderden mensen treft. Patiënten raken geleidelijk verlamd, doordat de motorische zenuwcellen in de hersenen en ruggenmerg afsterven. ALS is een progressieve, dodelijke ziekte, waarvoor geen goed medicijn is. De gemiddelde overleving na het ontstaan van symptomen is ongeveer 3 jaar. Een deel van de patiënten heeft echter een overleving van langer dan 10 jaar. Het enige medicijn dat het ziekteproces kan remmen is riluzole, maar het effect hiervan is beperkt.

Het ontstaan van de ziekte is grotendeels onbekend. De symptomen worden veroorzaakt door het afsterven van zenuwcellen, maar waarom dit gebeurt is onduidelijk. Bij een deel van de mensen met ALS (5-10%) is er sprake van een familiaire aandoening. Hierbij is er een ziekmakend gendefect dat binnen de familie wordt doorgegeven. Voor een deel (meer dan de helft) van de familiare ALS is inmiddels het oorzakelijke gendefect bekend. Hoe deze genetische afwijkingen leiden tot ziekte is echter onduidelijk.

Voor het grootste deel van ALS patiënten geldt echter dat de ziekte niet voorkomt in de familie. Deze sporadische vorm is waarschijnlijk multifactorieel bepaald: er moet een (genetische) aanleg zijn, mogelijk in combinatie met omgevingsfactoren zoals bijvoorbeeld voeding, roken, en blootstelling aan bepaalde stoffen. De genetische aanleg is, anders dan bij de familiaire vorm, waarschijnlijk niet bepaald door één gendefect, maar een combinatie van meerdere genetische variaties. Iedere genvariant verhoogt het risico op het krijgen van ALS, maar is afzonderlijk niet genoeg om de ziekte te veroorzaken. Echter, meerdere risicovarianten bij elkaar kunnen zorgen dat het risico zodanig verhoogd wordt dat iemand ziek wordt.

Ziekten (of in bredere zin: eigenschappen) die worden bepaald door meerdere genetische varianten (en omgevingsfactoren) worden "complexe" ziekten of eigenschappen genoemd. Andere voorbeelden naast sporadische ALS zijn bijvoorbeeld lichaamslengte, cholesterolwaarden in het bloed, diabetes, auto-immuunziekten en neurologische ziekten zoals de ziekte van Parkinson en Alzheimer dementie.

Om tot effectieve behandeling van deze ziekten te komen, is kennis nodig over het ontstaan hiervan. Eén invalshoek om tot deze kennis te komen is te onderzoeken welke genen het risico op een bepaalde ziekte beïnvloeden: wanneer een bepaald gen dit risico beïnvloedt, dan kan kennis over de functies van dat gen inzicht geven over de biologische mechanismen die tot de ziekte leiden.

In dit proefschrift beschrijf ik verschillende studies met dit doel: op verschillende manieren hebben we onderzocht welke genetische varianten het risico verhogen op het krijgen van ALS, om zo meer inzicht te krijgen in de oorzakelijke mechanismen en uiteindelijk een betere behandeling mogelijk te maken.

Genoom-wijde associatie studies
In de eerste hoofdstukken (hoofdstukken 2, 3 en 4) worden "genoom-wijde associatie studies" (GWAS) beschreven. Dit zijn studies waarmee wordt gezocht naar genetische varianten (single nucleotide polymorphisms, SNPs) die vaker voorkomen bij patiënten dan bij gezonde controlepersonen, en die dus geassocieerd zijn met ALS. Sinds enkele jaren is het mogelijk om in korte tijd honderdduizenden van deze SNPs te typeren bij honderden tot duizenden personen. Dit gebeurt met "SNP arrays", speciale chips waarop DNA wordt geanalyseerd door middel van probes die elk polymorfisme typeren. Zodoende is het mogelijk om de meest voorkomende genetische variaties door het hele genoom ("genoom-wijd") te onderzoeken. Per polymorfisme wordt getest of deze variatie vaker bij patiënten voorkomt. Omdat er honderdduizenden tests worden gedaan, is er een groot risico op toevalsbevindingen (vals-positieven). Hierom worden strenge statistische grenzen aangehouden om te kunnen stellen dat een genetische variant daadwerkelijk geassocieerd is. Eén manier is om het significantieniveau aan te passen aan het aantal statistische testen dat wordt uitgevoerd (Bonferroni correctie). Daarnaast is replicatie van ontdekte associaties in onafhankelijke studie populaties een manier van validatie.

In **hoofdstuk 2** wordt een GWAS beschreven waarin we het DNA van 461 Nederlandse ALS patiënten en 450 gezonde controlepersonen hebben onderzocht met SNP arrays. In totaal zijn ruim 311,000 genetische varianten onderzocht. Na statistische analyse werden de 485 meest geassocieerde SNPs getest in een tweede populatie van Nederlanders en Belgen. Op basis van de resultaten uit deze groep werden 17 SNPs geselecteerd en getest in een derde populatie van 313 Zweedse patienten en 303 controlepersonen. Eén SNP (rs2306677) was significant geassocieerd na Bonferroni correctie ($p = 0.01$). Deze SNP is gelegen in het gen *ITPR2* (inositol 1,4,5-triphosphate receptor 2). ITPR2 is betrokken bij de calciumhuishouding in zenuwcellen en een verstoord calciummetabolisme kan leiden tot celdood. *ITPR2* is dus een plausibele biologische kandidaat als ALS gen.

In **hoofdstuk 3** wordt een tweede GWAS beschreven waarin we data van de eerste GWAS combineren met data uit een Amerikaanse studie. De SNPs die in beide studies geassocieerd leken hebben we vervolgens getest in een tweede populatie van Nederlanders, Belgen en Zweden. Eén SNP (rs10260404) was geassocieerd in alle

bestudeerde populaties, met uitzondering van de Belgische. De risicovariant van deze SNP is geassocieerd met een mild verhoogd risico op het krijgen van ALS. Deze SNP ligt in het gen *DPP6*, dat tot expressie komt in het zenuwstelsel. Hoe de biologische functie van *DPP6* kan leiden tot motorneuron dood is niet bekend.

In **hoofdstuk 4** wordt een grote GWAS beschreven van meer dan 19,000 patienten en controles uit 9 landen, waarin we opnieuw door middel van genoom-wijde analyse van SNPs zoeken naar geassocieerde SNPs. Het wordt steeds duidelijker dat ziekte-geassocieerde SNPs zeer kleine effecten hebben (wat overigens niets zegt over de biologische betekenis van zulke vondsten). Om varianten met kleine effecten te kunnen vinden in een GWAS is veel statistische power nodig, wat inhoudt dat de studiepopulatie erg groot moet zijn om een statistisch significant effect aan te tonen. Met deze GWAS konden we twee gebieden in het genoom aanwijzen die geassocieerd zijn met een verhoogd risico op ALS: een gebied van gen *UNC13A* op chromosoom 19 en een gebied op chromosoom 9, met 3 genen. Hoe het gebied van *UNC13A* zorgt voor dood van motorneuronen is onbekend en moet verder worden bestudeerd. Uit later verschenen studies blijkt dat het gebied op chromosoom 9 een pathogene mutatie te bevatten in het gen *C9ORF72*, die waarschijnlijk de associatie verklaart. De onderliggende mechanismen die uiteindelijk leiden tot ALS worden momenteel onderzocht.

Copy number variation
Het volgende gedeelte in dit proefschrift omvat studies met als onderwerp "copy number variation". Copy number variants (CNVs) zijn deleties (minder kopieën vergeleken met een referentiegenoom) en inserties (meer kopieën vergeleken met een referentiegenoom) van DNA segmenten. Dit fenomeen werd vroeger vooral in verband gebracht met ernstige aangeboren afwijkingen, maar uit recenter onderzoek blijkt dat alle (gezonde) mensen duizenden van deze varianten bezitten. Deze kunnen net zoals SNPs geassocieerd zijn met (complexe) ziekten en andere fenotypen. Zo zijn bijvoorbeeld CNVs ontdekt die geassocieerd zijn met lichaamsgewicht, schizofrenie en auto-immuunziekten. In **hoofdstuk 5** analyseren we de genoom-wijde data uit eerdere GWAS studies op een andere manier dan in de SNP associatie studies: door middel van een zelf ontworpen algoritme zoeken we naar verschillen in hybridisatie intensiteit van de SNP probes. Als een probe gelokaliseerd is in een DNA segment dat is gedeleteerd in een bepaalde persoon, dan zal dit een minder sterk signaal geven bij analyse van de SNP array. Omgekeerd zal een duplicatie een sterker signaal geven. Door deze verschillen is het mogelijk om CNVs te detecteren in genoom-wijde SNP data. In de beschreven studie beschrijven we voor het eerst de karakteristieken van CNVs die voorkomen bij ALS patiënten: grootte, lokalisatie en welke genen door de CNVs worden beïnvloed. Tevens doen we een associatieanalyse ter identificatie van CNVs die vaker voorkomen bij patiënten (vergeleken met gezonde controles). De belangrijkste bevinding is dat er veel verschillende CNVs voorkomen, die elk erg zeldzaam zijn en meestal maar bij één individu voorkomen. De genen die gedeleteerd zijn bij ALS patiënten lijken als groep te verschillen van die in controlepersonen, maar het is niet mogelijk om individuele genen aan te wijzen die samenhangen met ALS pathogenese.

In **hoofdstuk 6** vergelijken we de bevindingen uit hoofdstuk 5 met data van Ierse patiënten en controles. We beschrijven 39 CNVs die in meerdere ALS patiënten voorkomen, maar niet in controlepersonen of in meer dan 10,000 personen in een CNV database. De betekenis hiervan is niet duidelijk: dit zijn ofwel zeldzame varianten

zonder relatie met ALS pathogenese, of het zijn zeldzame ALS-geassocieerde varianten. De zeldzaamheid van de varianten maakt het echter niet mogelijk om hier een uitspraak over te doen.

Om individuele ALS-geassocieerde CNV's te identificeren hebben we een grote CNV associatiestudie gedaan die wordt beschreven in **hoofdstuk 7**. In samenwerking met verschillende andere onderzoeksgroepen hebben we DNA van meer dan 19,000 personen onderzocht uit Nederland, België, Duitsland, Verenigd Koninkrijk, VS, Polen, Ierland en Scandinavië. De studie bestond uit 2 fasen: in de eerste fase hebben we genoom-wijd gezocht naar CNVs die geassocieerd leken te zijn met ALS en vervolgens hebben we twee genen (*NIPA1* en *DPP6)* verder onderzocht in een tweede studiepopulatie. In deze replicatiefase kwamen deleties in *NIPA1* en CNVs (duplicaties en deleties) in *DPP6* vaker voor bij patiënten (p=0.041 en p=0.097), maar dit voldeed niet aan onze eisen voor significantie. We concluderen dat deze data suggestief zijn voor mogelijke betrokkenheid van *NIPA1* en *DPP6* bij ALS, maar dat er geen formeel bewijs van associatie is. Overigens was *DPP6* al uit een eerdere GWAS naar voren gekomen en is *NIPA1* een gen dat is betrokken bij een andere neurologische ziekte (HSP 6), die overeenkomsten vertoont met ALS. Beide genen zijn dus interessante kandidaten voor verder onderzoek (zie ook hoofdstuk 9).

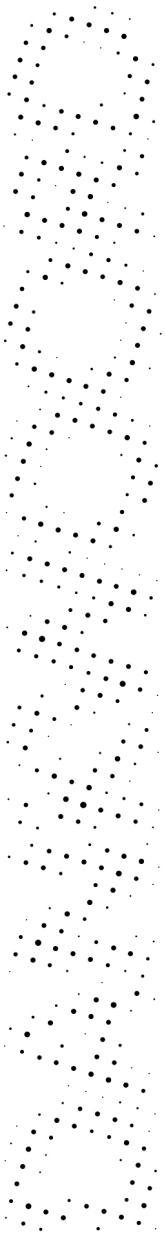
In **hoofdstuk 8** hebben we met een andere methode een associatiestudie gedaan met veelvoorkomende CNVs ("copy number polymorphisms", CNPs). Deze CNVs komen frequent voor in de populatie en worden met de eerder genoemde CNV studies door technische redenen niet goed onderzocht. Een deel van deze CNPs wordt indirect ook onderzocht in de conventionele SNP GWAS, maar een groot deel niet. Dit vormt dus een groep genetische varianten die worden gemist in eerdere studies. In deze studie gebruiken we opnieuw intensiteit data van SNP array hybridisaties, maar we gebruiken een andere methode om deze te analyseren. We testten een groep bekende CNPs in patienten en controles en keken per CNP of deze geassocieerd was met ALS en duur van overleving. Vijf CNPs die de sterkste associatie vertoonden hebben we vervolgens in een replicatieexperiment in een tweede populatie getest. Geen van deze CNPs bleek echter significant geassocieerd.

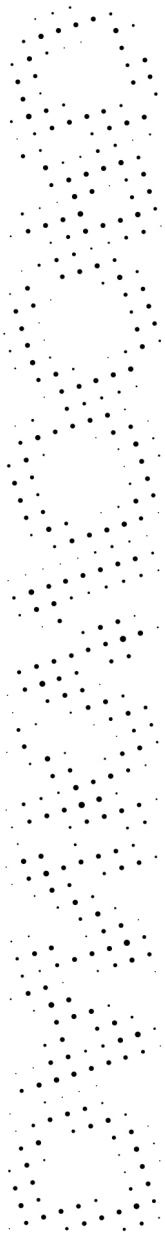
In **hoofdstuk 9** hebben we de rol van *NIPA1* bij ALS verder onderzocht. Zoals in hoofdstuk 7 beschreven, lijken *NIPA1* deleties een verhoogd risico op ALS te geven. In deze studie onderzochten we of (punt)mutaties in *NIPA1* geassocieerd zijn met ALS. Dit deden we door alle exonen van *NIPA1* te sequencen in een grote groep patiënten en controles (2,292 patiënten en 2,777 controles). In onze studiepopulatie vonden we in totaal twee schadelijke mutaties, beiden bij ALS patiënten. In controlepersonen vonden we geen schadelijke mutaties. Of deze mutaties daadwerkelijk iets met ALS te maken hebben, valt door de kleine aantallen niet te zeggen.

Verder onderzochten we een polyalanine repeat in *NIPA1*. Variaties in lengte van dit soort repeats in andere genen zijn geassocieerd met ziekte. De repeat in *NIPA1* was nooit eerder onderzocht in het kader van ziekte. Uit de associatietests bleek dat een langere polyalanine repeat is geassocieerd met zowel het risico op ALS, als de leeftijd waarop de symptomen ontstaan en de overleving. Mensen met een lange *NIPA1* repeat hebben een licht verhoogd risico op het krijgen van ALS (odds ratio = 1,71). Patiënten met een lange repeat worden ongeveer 3,6 jaar eerder ziek en overlijden ongeveer 3 maanden eerder. Samen met de eerdere bevinding dat *NIPA1* deleties vaker voorkomen bij ALS

patiënten, wijzen deze bevindingen op een rol van *NIPA1* bij de pathogenese van ALS. Hoe genetische varianten in *NIPA1* leiden tot ziekte moet verder worden onderzocht. In **hoofdstuk 10** beschrijf ik een studie naar de betrokkenheid van het *SMN* gen bij ALS. Het *SMN* (Survival of Motor Neuron) gen is het causale gen bij de ziekte Spinale Spieratrofie (SMA), een aangeboren motor neuron ziekte bij kinderen. Omdat deze ziekte overeenkomsten vertoont met ALS is al eerder onderzocht of *SMN* ook een rol speelt in het ontstaan van ALS. Uit deze eerdere studies kwamen geen eenduidige resultaten, zodat de precieze rol van *SMN* onduidelijk bleef. Dit was deels te wijten aan het feit dat dit gen technisch lastig te onderzoeken is en dat eerdere studies relatief klein waren. Om een definitief antwoord te kunnen geven op de vraag of genetische varianten in *SMN* ook geassocieerd zijn met ALS hebben we een grote associatiestudie gedaan met een relatief nieuwe techniek (MLPA). De belangrijkste bevinding is dat duplicaties van *SMN1* een verhoogd risico geven op ALS. Uit een meta-analyse van eerdere data lijkt dit effect stand te houden. In tegenstelling tot resultaten van eerdere studies waren deleties van *SMN1* en variaties in *SMN2* (een *SMN1* homoloog) niet geassocieerd met ALS. Het is niet duidelijk hoe het is te verklaren dat duplicaties van *SMN1*, maar niet deleties, zijn geassocieerd met ALS. Dit zal verder moeten worden onderzocht.

In **hoofdstuk 11** vat ik de beschreven studies samen en plaats ik deze in een breder kader.





Dit proefschrift is mogelijk gemaakt door de directe en indirecte steun van velen. Ik zal er niet in slagen om iedereen hier recht te doen, maar ik waag een poging om een aantal mensen te bedanken.

Professor van den Berg, beste Leonard. Het is een groot geluk geweest om in jouw groep terecht te komen. Je bent er in korte tijd in geslaagd om een groeiende en bloeiende onderzoeksgroep op te zetten, met enthousiasme en humor die op iedereen aanstekelijk werken. Met jouw gevoel voor relativering maak je complexe zaken prettig simpel (en simpele zaken soms prettig complex). Je hebt gelijk: eigenlijk draait dit proefschrift vooral om "stipjes". Ik hoop dat er in de toekomst gelegenheid blijft om samen te werken.

Dr Veldink, beste Jan. Ik heb zelden snellere denkers ontmoet dan jij. Regelmatig kostte het mij enige tijd en mentale arbeid om jouw adviezen te doorgronden, maar deze bleken altijd van grote waarde. Zonder jouw hulp had dit proefschrift (en dat van vele anderen) er heel pover uitgezien. Ik heb genoten van de samenwerking!

Professor Ophoff, beste Roel. Mede dankzij jou hebben we met het genetische onderzoek binnen onze groep een grote stap voorwaarts kunnen zetten. Bedankt voor de samenwerking!

Chris en Michael. Als "ALS boys" op het lab vormden we een mooi trio, zowel wat betreft ons onderzoek als daarbuiten. Michael, ik bewonder jouw humor en effectiviteit waarvan ik, Captain Slow, nog veel kan leren. Chris, zelfs onder grote tijdsdruk (bijvoorbeeld na een veelbetekenend "Succes!" van Leonard) was jij altijd bereid om te helpen door in alle rust niet-werkende scripts te repareren, prachtige plots te produceren, enzovoort. Ik heb genoten van jullie gezelschap, veel dank voor de mooie jaren!

Beste Pawl. Jij stond aan de wieg van het genetische onderzoek binnen onze groep. Ik weet nog goed dat er in den beginne wat lacherig werd gedaan over de "snipjes en snapjes", maar deze zijn uiteindelijk niet onbelangrijk geweest, toch? Je nam altijd de tijd om mij en vele anderen te helpen met labzaken waarvan wij, domme dokters, geen verstand hadden, waarvoor heel veel dank.

Ewout, Elies, Sanne, Esther, Wouter, Frank, Marka, Max. Dank voor alle gezelligheid en het bewaken van de klok (10.00, 11.30, 15.00), zodat koffie en lunch op precies de juiste tijd werden geconsumeerd!

Henk, Peter, Eric, Ruben. Hoeveel proefschriften zijn in de loop der jaren niet tot stand gekomen dank zij jullie hulp? Zo ook het mijne. Veel dank!

Pim, Mark, Henk en Ruben van het Hubrecht: dank voor alle hulp bij het sequencen en de lol op jullie lab. Wat hebben jullie het goed begrepen: hard werken en gezelligheid gaan goed samen!

Nienke en Inge, waar zou de MND poli zijn zonder jullie? Door jullie betrokkenheid was het daar ondanks alle ellende toch altijd een fijne plek om te werken.

Lude, jouw geestdrift en originaliteit inspireerden mij altijd enorm. Wat een voorrecht om met jou te mogen werken!

Ewoud, zonder jouw hulp zou ik nu nog steeds aan het prutsen zijn om BLAS (wat?) aan de praat te krijgen. Hoewel onze fietstripjes de laatste jaren iets minder heldhaftig geworden zijn (onder de 10 graden niet meer op de fiets, niet meer boven de 200 km per dag...), kijk ik uit naar onze toekomstige rondjes. Binnenkort maar weer eens de heuvels opzoeken?

Ilse, hoewel het boekje nu nog niet eens af is, ben ik er door jouw vakmanschap al van overtuigd dat het prachtig gaat worden. Dank voor de prettige samenwerking! (En sorry voor het telkens oprekken van de door mij zelf gestelde deadlines...).

Professor Wokke en Professor van Gijn. Dank voor het vertrouwen om mij aan te nemen voor de opleiding tot neuroloog. Het is mij een grote eer en plezier om in onze kliniek te worden opgeleid!

Mijn collega's van de afdeling neurologie van het UMC Utrecht. Dankzij jullie voel ik mij inmiddels net zo op mijn plaats in de kliniek als in het lab. Ik vind het een voorrecht om deel van jullie uit te mogen maken.

Maarten, Sebas, Daan, David, Jorn, Joost, Sven, Maurits, Wouter, Harm, Jan Pieter, Jelte, Reinier, Stoops, Marinus, Rutger, Diederik. Ik ken weinig kleurrijker gezelschappen dan jullie. Bij jullie is het altijd weer moeilijk om de tranen van het lachen te bedwingen. Ik hoop dat jullie geen voorbeeld aan mij nemen en dat we elkaar zien op 18 of 21 december!

Mijn paranimfen Jurre en Michael, wat een rustgevende en gezellige gedachte dat jullie mij bijstaan!

Ger en Wil. Dank voor jullie schoonouderlijke betrokkenheid en warmte!

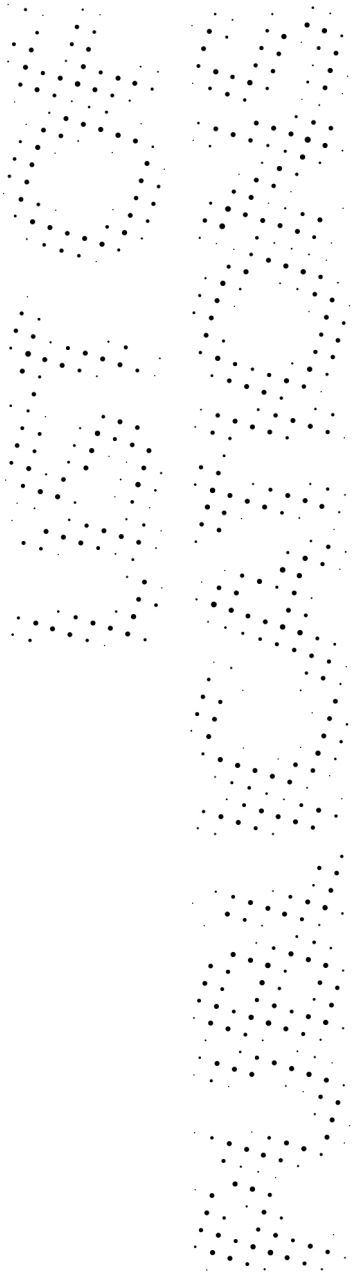
Mijn lieve Geschwistern. Jurre, ruwe bolster, blanke pit, ik bewonder jouw enorme originaliteit en eigenwijsheid, met een humor waar weinigen tegen bestand zijn. I like your style, dude! Lieve Sanne, wat ben je toch een intens fijn mens! Jouw onverwoestbaar goede humeur is een verademing aan het eind van een koude, natte, blubberige dag in Northumbria. Ondanks drie volle agenda's hoop ik dat het ons lukt om snel weer met elkaar op pad te gaan. Toch maar eens de GR 20 proberen?

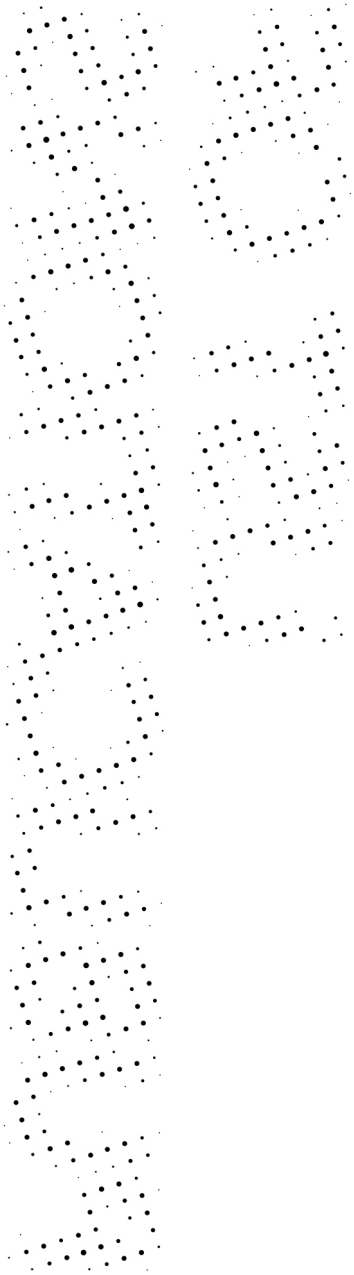
Lieve Pa en Ma. Aan jullie heb ik alles te danken. Dankzij jullie hebben Jurre, Sanne en ik een heerlijke jeugd in Zeeland gehad (en moet Marieke nu al mijn nostalgische praat aanhoren...). Door jullie ongedwongenheid waren wij altijd vrij om ons eigen plan te trekken, altijd door jullie gesteund. Ik vind het fantastisch om jullie te zien genieten van de rol van pake en oma van Mies. Dat jullie maar vaak langs mogen blijven komen!

Lieve Mies, mijn prachtige dochter, wat ben ik trots op je! Gelukkig ben jij er om mij er telkens op te wijzen wat echt belangrijk is!

Lieve Marieke, mijn Beste! Wat ben ik blij dat we samen zijn, hic anda thu! Dat het inderdaad maar altijd zo mag blijven. Wil je nooit met me trouwen?

Hylke.





THIS THESIS

Blauw, H.M., Barnes, C.P., van Vught, P.W., van Rheenen, W., Verheul, M., Cuppen, E., Veldink, J.H. and van den Berg, L.H. (2012) *SMN1* gene duplications are associated with sporadic ALS. *Neurology*, 78, 776-780.

Blauw, H.M.[#], Van Rheenen, W.[#], Koppers, M., Van Damme, P., Waibel, S., Lemmens, R., Van Vught, P.W.J., Meyer, T., Schulte, C., Gasser, T., Cuppen, E., Pasterkamp, R.J., Robberecht, W., Ludolph, A.C., Veldink, J.H. and Van Den Berg, L.H. (2012) *NIPA1* polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. *Human Molecular Genetics*, 21, 2497-2502. ([#] joint first authors)

Blauw, H.M., Al-Chalabi, A., Andersen, P.M., van Vught, P.W., Diekstra, F.P., van Es, M.A., Saris, C.G., Groen, E.J., van Rheenen, W., Koppers, M., Van't Slot, R., Strengman, E., Estrada, K., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Kiemeney, L.A., Vermeulen, S.H., Birve, A., Waibel, S., Meyer, T., Cronin, S., McLaughlin, R.L., Hardiman, O., Sapp, P.C., Tobin, M.D., Wain, L.V., Tomik, B., Slowik, A., Lemmens, R., Rujescu, D., Schulte, C., Gasser, T., Brown, R.H., Jr., Landers, J.E., Robberecht, W., Ludolph, A.C., Ophoff, R.A., Veldink, J.H. and van den Berg, L.H. (2010) A large genome scan for rare CNVs in amyotrophic lateral sclerosis. *Human Molecular Genetics*, 19, 4091-4099.

Van Es, M.A., Veldink, J.H., Saris, C.G.J., **Blauw, H.M.**, Van Vught, P.W.J., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J.N., Huisman, M.H.B., Van Der Kooi, A.J., De Visser, M., Dahlberg, C., Estrada, K., Rivadeneira, F., Hofman, A., Zwarts, M.J., Van Doormaal, P.T.C., Rujescu, D., Strengman, E., Giegling, I., Muglia, P., Tomik, B., Slowik, A., Uitterlinden, A.G., Hendrich, C., Waibel, S., Meyer, T., Ludolph, A.C., Glass, J.D., Purcell, S., Cichon, S., Nöthen, M.M., Wichmann, H.-E., Schreiber, S., Vermeulen, S.H.H.M., Kiemeney, L.A., Wokke, J.H.J., Cronin, S., Mclaughlin, R.L., Hardiman, O., Fumoto, K., Pasterkamp, R.J., Meininger, V., Melki, J., Leigh, P.N., Shaw, C.E., Landers, J.E., Al-Chalabi, A., Brown, R.H., Robberecht, W., Andersen, P.M., Ophoff, R.A. and Van Den Berg, L.H. (2009) Genome-wide association study identifies 19p13.3 (*UNC13A*) and 9p21.2 as susceptibility loci for

sporadic amyotrophic lateral sclerosis. *Nature Genetics*, 41, 1083-1087.

Blauw, H.M.[#], Cronin, S.[#], Veldink, J.H., van Es, M.A., Ophoff, R.A., Bradley, D.G., van den Berg, L.H. and Hardiman, O. (2008) Analysis of genome-wide copy number variation in Irish and Dutch ALS populations. *Human Molecular Genetics*, 17, 3392-3398. ([#] joint first authors)

Blauw, H.M.[#], van Es, M.A.[#], van Vught, P.W.[#], Franke, L.[#], Saris, C.G., Van den Bosch, L., de Jong, S.W., de Jong, V., Baas, F., van't Slot, R., Lemmens, R., Schelhaas, H.J., Birve, A., Sleegers, K., Van Broeckhoven, C., Schymick, J.C., Traynor, B.J., Wokke, J.H., Wijmenga, C., Robberecht, W., Andersen, P.M., Veldink, J.H., Ophoff, R.A. and van den Berg, L.H. (2008) Genetic variation in *DPP6* is associated with susceptibility to amyotrophic lateral sclerosis. *Nature Genetics*, 40, 29-31. ([#] joint first authors)

Blauw, H.M.[#], Veldink, J.H.[#], Van Es, M.A., Van Vught, P.W., Saris, C.G.J., van der Zwaag, B., Franke, L., Burbach, J.P.H., Wokke, J.H., Ophoff, R.A. and van den Berg, L.H. (2008) Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen. *Lancet Neurology*, 7, 319-326. ([#] joint first authors)

Blauw, H.M.[#], Van Es, M.A.[#], Van Vught, P.W.[#], Franke, L.[#], Saris, C.G., Andersen, P.M., Van Den Bosch, L., De Jong, S.W., van 't Slot, R., Birve, A., Lemmens, R., de Jong, V., Baas, F., Schelhaas, H.J., Sleegers, K., Van Broeckhoven, C., Wokke, J.H.J., Wijmenga, C., Robberecht, W., Veldink, J.H., Ophoff, R.A. and Van Den Berg, L.H. (2007) *ITPR2* as a susceptibility gene in sporadic amyotrophic lateral sclerosis: a genome-wide association study. *Lancet Neurology*, 6, 869-877. ([#] joint first authors)

OTHER

The ALSGEN Consortium (2012) Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. Neurobiology of Aging, Epub ahead of print.

Diekstra, F.P., Saris, C.G., van Rheenen, W., Franke, L., Jansen, R.C., van Es, M.A., van Vught, P.W., **Blauw, H.M.**, Groen, E.J., Horvath, S., Estrada, K., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Robberecht, W., Andersen, P.M., Melki, J., Meininger, V., Hardiman, O., Landers, J.E., Brown, R.H., Jr., Shatunov, A., Shaw, C.E., Leigh, P.N., Al-Chalabi, A., Ophoff, R.A., van den Berg, L.H. and Veldink, J.H. (2012) Mapping of gene expression reveals *CYP27A1* as a susceptibility gene for sporadic ALS. *PLoS One*, 7, e35333.

Van Es, M.A., Schelhaas, H.J., Van Vught, P.W.J., Ticozzi, N., Andersen, P.M., Groen, E.J.N., Schulte, C., **Blauw, H.M.**, Koppers, M., Diekstra, F.P., Fumoto, K., Leclerc, A.L., Keagle, P., Bloem, B.R., Scheffer, H., Van Nuenen, B.F.L., Van Blitterswijk, M., Van Rheenen, W., Wills, A.-M., Lowe, P.P., Hu, G.-F., Yu, W., Kishikawa, H., Wu, D., Folkerth, R.D., Mariani, C., Goldwurm, S., Pezzoli, G., Van Damme, P., Lemmens, R., Dahlberg, C., Birve, A., Fernández-Santiago, R., Waibel, S., Klein, C., Weber, M., Van Der Kooi, A.J., De Visser, M., Verbaan, D., Van Hilten, J.J., Heutink, P., Hennekam, E.A.M., Cuppen, E., Berg, D., Brown, R.H., Silani, V., Gasser, T., Ludolph, A.C., Robberecht, W., Ophoff, R.A., Veldink, J.H., Pasterkamp, R.J., De Bakker, P.I.W., Landers, J.E., Van De Warrenburg, B.P. and van den Berg, L.H. (2011) Angiogenin variants in Parkinson disease and amyotrophic lateral sclerosis. *Annals of Neurology.*, 70, 964-973.

Knauff, E.A., **Blauw, H.M.**, Pearson, P.L., Kok, K., Wijmenga, C., Veldink, J.H., van den Berg, L.H., Bouchard, P., Fauser, B.C. and Franke, L. (2011) Copy number variants on the X chromosome in women with primary ovarian insufficiency. *Fertility and Sterility*, 95, 1584-1588 e1581.

van Es, M.A., van Vught, P.W.J., van Kempen, G., **Blauw, H.M.**, Veldink, J.H. and Van den Berg, L.H. (2009) *DPP6* is associated with susceptibility to progressive spinal muscular atrophy. *Neurology*, 72, 1184-1185.

Landers, J.E., Melki, J., Meininger, V., Glass, J.D., van den Berg, L.H., Van Es, M.A., Sapp, P.C., Van Vught, P.W.J., McKenna-Yasek, D.M., **Blauw, H.M.**, Cho, T.-J., Polak, M., Shi, L., Wills, A.-M., Broom, W.J., Ticozzi, N., Silani, V., Ozoguz, A., Rodriguez-Leyva, I., Veldink, J.H., Ivinson, A.J., Saris, C.G.J., Hosler, B.A., Barnes-Nessa, A., Couture, N., Wokke, J.H.J., Kwiatkowski, T.J., Ophoff, R.A., Cronin, S., Hardiman, O., Diekstra, F.P., Leigh, P.N., Shaw, C.E., Simpson, C.L., Hansen, V.K., Powell, J.F., Corcia, P., Salachas, F., Heath, S., Galan, P., Georges, F., Horvitz, H.R., Lathrop, M., Purcell, S., Al-Chalabi, A. and Brown, R.H. (2009) Reduced expression of the Kinesin-Associated Protein 3 (*KIFAP3*) gene increases survival in sporadic amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences USA*, 106, 9004-9009.

Wang, H., Veldink, J.H., **Blauw, H.M**., van den Berg, L.H., Ophoff, R.A. and Sabatti, C. (2009) Markov Models for inferring copy number variations from genotype data on Illumina platforms. *Human Heredity*, 68, 1-22.

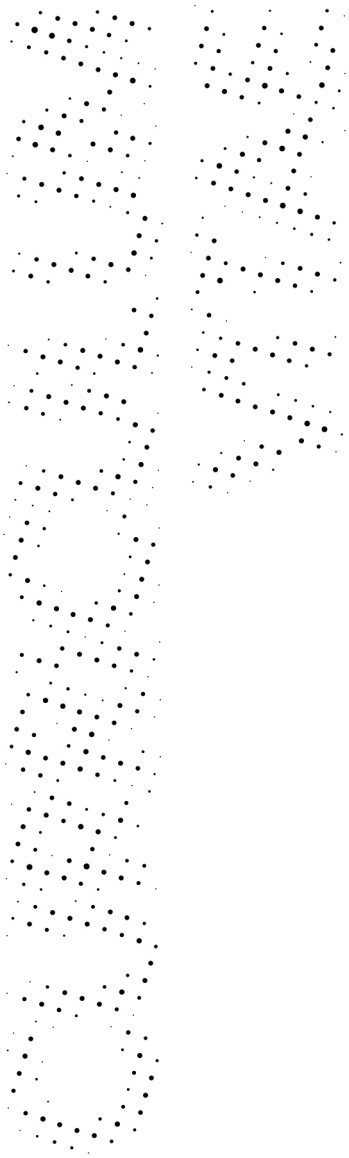
van der Zwaag, B., Franke, L., Poot, M., Hochstenbach, R., Spierenburg, H.A., Vorstman, J.A., van Daalen, E., de Jonge, M.V., Verbeek, N.E., Brilstra, E.H., van 't Slot, R., Ophoff, R.A., van Es, M.A., **Blauw, H.M**., Veldink, J.H., Buizer-Voskamp, J.E., Beemer, F.A., van den Berg, L.H., Wijmenga, C., van Amstel, H.K., van Engeland, H., Burbach, J.P. and Staal, W.G. (2009) Gene-network analysis identifies susceptibility genes related to glycobiology in autism. *PLoS One*, 4, e5324.

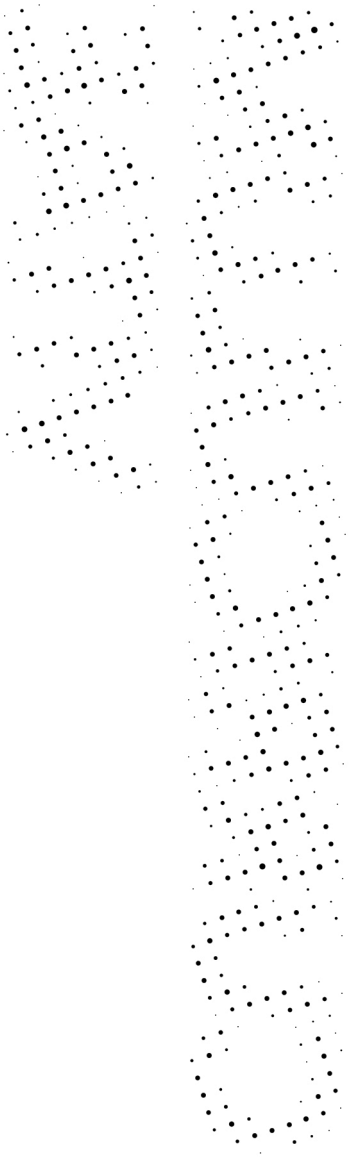
Saris, C.G., Horvath, S., van Vught, P.W., van Es, M.A., **Blauw, H.M**., Fuller, T.F., Langfelder, P., DeYoung, J., Wokke, J.H., Veldink, J.H., van den Berg, L.H. and Ophoff, R.A. (2009) Weighted gene co-expression network analysis of the peripheral blood from Amyotrophic Lateral Sclerosis patients. *BMC Genomics*, 10, 405.

Franke, L., de Kovel, C.G., Aulchenko, Y.S., Trynka, G., Zhernakova, A., Hunt, K.A., **Blauw, H.M**., van den Berg, L.H., Ophoff, R., Deloukas, P., van Heel, D.A. and Wijmenga, C. (2008) Detection, imputation, and association analysis of small deletions and null alleles on oligonucleotide arrays. *American Journal of Human Genetics*, 82, 1316-1333.

Mager, J.J., Overtoom, T.T., **Blauw, H.M**., Lammers, J.W. and Westermann, C.J. (2004) Embolotherapy of pulmonary arteriovenous malformations: long-term results in 112 patients. *Journal of Vascular and Interventional Radiology*, 15, 451-456.

Groeneveld, G.J., van Muiswinkel, F.L., de Leeuw van Weenen, J., **Blauw, H.M**., Veldink, J.H., Wokke, J.H., van den Berg, L.H. and Bar, P.R. (2004) CGP 3466B has no effect on disease course of (G93A) mSOD1 transgenic mice. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders*, 5, 220-225.





Hylke Merijn Blauw werd geboren in Eutin (Duitsland) op 26 juni 1979. Na het voltooien van het VWO op de Christelijke Scholengemeenschap Walcheren in Middelburg studeerde hij geneeskunde aan de Universiteit Utrecht. Tijdens zijn studie deed hij wetenschappelijk onderzoek in het Laboratorium voor Experimentele Neurologie in het UMC Utrecht, onder supervisie van dr. G.J. Groenveld en prof. dr. P.R. Bär. Na zijn artsexamen in 2006 begon hij in hetzelfde lab met het onderzoek dat resulteerde in dit proefschrift, met als promotor prof. dr L.H. van den Berg. Tijdens deze periode begon hij ook met zijn opleiding tot neuroloog in het UMC Utrecht.