Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis

Wouter van Rheenen[1,123], Aleksey Shatunov[2,123], Annelot M. Dekker[1], Russell L. McLaughlin[3], Frank P. Diekstra[1], Sara L. Pulit[4], Rick A.A. van der Spek[1], Urmo Võsa[5], Simone de Jong[6,7], Matthew R. Robinson[8], Jian Yang[8], Isabella Fogh[2,9], Perry T.C. van Doormaal[1], Gijs H.P. Tazelaar[1], Max Koppers[1,10], Anna M. Blokhuis[1,10], William Sproviero[2], Ashley R. Jones[2], Kevin P. Kenna[11], Kristel R. van Eijk[1], Oliver Harschnitz[1,10], Raymond D. Schellevis[1], William J. Brands[1], Jelena Medic[1], Androniki Menelaou[4], Alice Vajda[12,13], Nicola Ticozzi[9,14], Kuang Lin[2], Boris Rogelj[15,16], Katarina Vrabec[17], Metka Ravnik-Glavač[17,18], Blaž Koritnik[19], Janez Zidar[19], Lea Leonardis[19], Leja Dolenc Grošelj[19], Stéphanie Millecamps[20], François Salachas[20,21,22], Vincent Meininger[23,24], Mamede de Carvalho[25,26], Susana Pinto[25,26], Jesus S. Mora[27], Ricardo Rojas-García[28,29], Meraida Polak[30,31], Siddharthan Chandran[32,33], Shuna Colville[32], Robert Swingler[32], Karen E. Morrison[34], Pamela J. Shaw[35], John Hardy[36], Richard W. Orrell[37], Alan Pittman[36,38], Katie Sidle[37], Pietro Fratta[39], Andrea Malaspina[40,41], Simon Topp[2], Susanne Petri[42], Susanne Abdulla[43], Carsten Drepper[44], Michael Sendtner[44], Thomas Meyer[45], Roel A. Ophoff[46,47,48], Kim A. Staats[48], Martina Wiedau-Pazos[49], Catherine Lomen-Hoerth[50], Vivianna M. Van Deerlin[51], John Q. Trojanowski[51], Lauren Elman[52], Leo McCluskey[52], A. Nazli Basak[53], Ceren Tunca[53], Hamid Hamzeiy[53], Yesim Parman[54], Thomas Meitinger[55], Peter Lichtner[55], Milena Radivojkov-Blagojevic[55], Christian R. Andres[56], Cindy Maurel[56], Gilbert Bensimon[57,58,59], Bernhard Landwehrmeyer[60], Alexis Brice[61,62,63,64,65], Christine A.M. Payan[57,59], Safaa Saker-Delye[66], Alexandra Dürr[67], Nicholas W. Wood[68], Lukas Tittmann[69], Wolfgang Lieb[69], Andre Franke[70], Marcella Rietschel[71], Sven Cichon[72,73,74,75,76], Markus M. Nöthen[72,73], Philippe Amouyel[77], Christophe Tzourio[78], Jean-François Dartigues[78], Andre G. Uitterlinden[79,80], Fernando Rivadeneira[79,80], Karol Estrada[79], Albert Hofman[80,81], Charles Curtis[6,7], Hylke M. Blauw[1], Anneke J. van der Kooi[82], Marianne de Visser[82], An Goris[83], Markus Weber[84], Christopher E. Shaw[2], Bradley N. Smith[2], Orietta Pansarasa[85], Cristina Cereda[85], Roberto Del Bo[86], Giacomo P. Comi[86], Sandra D'Alfonso[87], Cinzia Bertolin[88], Gianni Sorarù[88], Letizia Mazzini[89], Viviana Pensato[90], Cinzia Gellera[90], Cinzia Tiloca[9], Antonia Ratti[9,14], Andrea Calvo[91,92], Cristina Moglia[91,92], Maura Brunetti[91,92], Simona Arcuti[93], Rosa Capozzo[93], Chiara Zecca[93], Christian Lunetta[94], Silvana Penco[95], Nilo Riva[96], Alessandro Padovani[97], Massimiliano Filosto[97], Bernard Muller[98], Robbert Jan Stuit[98], PARALS registry[99], SLALOM group[99], SLAP registry[99], FALS Sequencing Consortium[99],

SLAGEN Consortium[99], NNIPPS Study Group[99], Ian Blair[100], Katharine Zhang[100], Emily P. McCann[100], Jennifer A. Fifita[100], Garth A. Nicholson[100,101], Dominic B. Rowe[100], Roger Pamphlett[102], Matthew C. Kiernan[103], Julian Grosskreutz[104], Otto W. Witte[104], Thomas Ringer[104], Tino Prell[104], Beatrice Stubendorff[104], Ingo Kurth[105], Christian A. Hübner[105], P. Nigel Leigh[106], Federico Casale[91], Adriano Chio[91,92], Ettore Beghi[107], Elisabetta Pupillo[107], Rosanna Tortelli[93], Giancarlo Logroscino[108,109], John Powell[2], Albert C. Ludolph[60], Jochen H. Weishaupt[60], Wim Robberecht[83,110,111], Philip Van Damme[83,110,111], Lude Franke[5], Tune H. Pers[112,113,114,115,116], Robert H. Brown[11], Jonathan D. Glass[30,31], John E. Landers[11], Orla Hardiman[12,13], Peter M. Andersen[60,117], Philippe Corcia[56,118,119], Patrick Vourc'h[56], Vincenzo Silani[9,14], Naomi R. Wray[8], Peter M. Visscher[8,120], Paul I.W. de Bakker[4,121], Michael A. van Es[1], R. Jeroen Pasterkamp[10], Cathryn M. Lewis[6,122], Gerome Breen[6,7], Ammar Al-Chalabi[2,124,125], Leonard H. van den Berg[1,124] & Jan H. Veldink[1,124,125]

Affiliations:

1. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands.
2. Maurice Wohl Clinical Neuroscience Institute, King's College London, Department of Basic and Clinical Neuroscience, London, UK.
3. Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin, Dublin, Republic of Ireland.
4. Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands.
5. Department of Genetics, University of Groningen, University Medical Centre Groningen, Groningen, The Netherlands.
6. MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.
7. NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK.
8. Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia.
9. Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico Italiano, Milano, Italy.
10. Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands.

11. Department of Neurology, University of Massachusetts Medical School, Worcester, MA, USA.
12. Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences Institute, Dublin, Republic of Ireland.
13. Department of Neurology, Beaumont Hospital, Dublin, Republic of Ireland.
14. Department of Pathophysiology and Tranplantation, 'Dino Ferrari' Center, Università degli Studi di Milano, Milano, Italy.
15. Department of Biotechnology, Jožef Stefan Institute, Ljubljana, Slovenia.
16. Biomedical Research Institute BRIS, Ljubljana, Slovenia.
17. Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine, University of Ljubljana, SI-1000 Ljubljana, Slovenia.
18. Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, SI-1000 Ljubljana, Slovenia.
19. Ljubljana ALS Centre, Institute of Clinical Neurophysiology, University Medical Centre Ljubljana, SI-1000 Ljubljana, Slovenia.
20. Institut du Cerveau et de la Moelle épinière, Inserm U1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMRS1127, Paris, France.
21. Centre de Référence Maladies Rares SLA Ile de France, Département de Neurologie, Hôpital de la Pitié-Salpêtrière, Paris, France.
22. GRC-UPMC SLA et maladies du Motoneurone, France.
23. Ramsay Generale de Santé, Hopital Peupliers, Paris, France.
24. Réseau SLA Ile de France.
25. Institute of Physiology, Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon, Portugal.
26. Department of Neurosciences, Hospital de Santa Maria-CHLN, Lisbon, Portugal.
27. Hospital San Rafael, Madrid, Spain
28. Neurology Department, Hospital de la Santa Creu i Sant Pau de Barcelona, Autonomous University of Barcelona, Barcelona, Spain.
29. Centro de Investigación en red en Enfermedades Raras (CIBERER), Spain.
30. Department Neurology, Emory University School of Medicine, Atlanta, GA, USA.
31. Emory ALS Center, Emory University School of Medicine, Atlanta, GA, USA.
32. Euan MacDonald Centre for Motor Neurone Disease Research, Edinburgh, UK.
33. Centre for Neuroregeneration and Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK.

34. Faculty of Medicine, University of Southampton, Southampton, UK.
35. Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield, Sheffield, UK.
36. Department of Molecular Neuroscience, Institute of Neurology, University College London, UK.
37. Department of Clinical Neuroscience, Institute of Neurology, University College London, UK.
38. Reta Lila Weston Institute, Institute of Neurology, University College London, UK.
39. Sobell Department of Motor Neuroscience and Movement Disorders, Institute of Neurology, University College London
40. Centre for Neuroscience and Trauma, Blizard Institute, Queen Mary University of London, London, UK.
41. North-East London and Essex Regional Motor Neuron Disease Care Centre, London, UK.
42. Department of Neurology, Hannover Medical School, Hannover, Germany.
43. Department of Neurology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany.
44. Institute of Clinical Neurobiology, University Hospital Wuerzburg, Germany.
45. Charité University Hospital, Humboldt-University, Berlin, Germany.
46. University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, The Netherlands.
47. Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, CA, USA.
48. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA.
49. Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA.
50. Department of Neurology, University of California, San Francisco, CA, USA.
51. Center for Neurodegenerative Disease Research, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA.
52. Department of Neurology, Perelman School of Medicine at the University of Pennsylvania, PA USA.
53. Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey.
54. Neurology Department, Istanbul Medical School, Istanbul University, Istanbul, Turkey.
55. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.

56. INSERM U930, Université François Rabelais, Tours, France.
57. APHP, Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpêtrière, France.
58. Université Pierre & Marie Curie, Pharmacologie, Paris VI, Paris, France.
59. BESPIM, CHU-Nîmes, Nîmes, France.
60. Department of Neurology, Ulm University, Ulm, Germany.
61. INSERM U 1127, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
62. CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
63. Sorbonne Universités, Université Pierre et Marie Curie Paris 06 UMRS 1127, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
64. Institut du Cerveau et de la Moelle épinière, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
65. APHP, Département de Génétique, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
66. Genethon, CNRS UMR 8587 Evry, France.
67. Department of Medical Genetics, l'Institut du Cerveau et de la Moelle Épinière, Hoptial Salpêtrière, 75013 Paris, France.
68. Department of Neurogenetics, Institute of Neurology, University College London, UK.
69. PopGen Biobank and Institute of Epidemiology, Christian Albrechts-University Kiel, Kiel, Germany.
70. Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany.
71. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Faculty of Medicine Mannheim, University of Heidelberg, Germany
72. Institute of Human Genetics, University of Bonn, Bonn, Germany.
73. Department of Genomics, Life and Brain Center, Bonn, Germany.
74. Division of Medical Genetics, University Hospital Basel, University of Basel, Basel, Switzerland.
75. Department of Biomedicine, University of Basel, Basel, Switzerland.
76. Institute of Neuroscience and Medicine INM-1, Research Center Juelich, Juelich, Germany.
77. University of Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk Factor and molecular determinants of aging diseases, Labex Distalz, F-59000 Lille, France.
78. Bordeaux University, ISPED, Centre INSERM U1219-Epidemiologie-Biostatistique & CIC-1401, CHU de Bordeaux, Pole de Sante Publique, Bordeaux, France.
79. Department of Internal Medicine, Genetics Laboratory, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands.

80. Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, The Netherlands.
81. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.
82. Department of Neurology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.
83. KU Leuven - University of Leuven, Department of Neurosciences, Experimental Neurology and Leuven Research Institute for Neuroscience and Disease (LIND), B-3000 Leuven, Belgium
84. Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, 9007, St. Gallen, Switzerland.
85. Laboratory of Experimental Neurobiology, IRCCS 'C. Mondino' National Institute of Neurology Foundation, Pavia, Italy.
86. Neurologic Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.
87. Department of Health Sciences, Interdisciplinary Research Center of Autoimmune Diseases, UPO, Università del Piemonte Orientale, Novara, Italy.
88. Department of Neurosciences, University of Padova, Padova, Italy.
89. Department of Neurology, University of Eastern Piedmont, Novara, Italy.
90. Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico 'Carlo Besta', Milano, Italy.
91. "Rita Levi Montalcini" Department of Neuroscience, ALS Centre, University of Torino, Turin, Italy.
92. Azienda Ospedaliera Città della Salute e della Scienza, Torino, Italy.
93. Department of Clinical research in Neurology, University of Bari "A. Moro", at Pia Fondazione "Card. G. Panico", Tricase (LE), Italy.
94. NEMO Clinical Center, Serena Onlus Foundation, Niguarda Ca' Granda Hostipal, Milan, Italy.
95. Medical Genetics Unit, Department of Laboratory Medicine, Niguarda Ca' Granda Hospital, Milan, Italy.
96. Department of Neurology, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy.
97. Neurology Unit, Department of Clinical and Experimental Sciences, University of Brescia, Italy.

98. Project MinE Foundation, Rotterdam, The Netherlands.
99. A list of members and affiliations appears in the Supplementary Note.
100. Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Macquarie University, Sydney, New South Wales, Australia.
101. University of Sydney, ANZAC Research Institute, Concord Hospital, Sydney, New South Wales, Australia.
102. The Stacey MND Laboratory, Department of Pathology, The University of Sydney, Sydney, New South Wales, Australia.
103. Brain and Mind Centre, The University of Sydney, New South Wales 2050, Australia.
104. Hans-Berger Department of Neurology, Jena University Hospital, Jena, Germany.
105. Institute of Human Genetics, Jena University Hospital, Jena, Germany.
106. Department of Neurology, Brighton and Sussex Medical School Trafford Centre for Biomedical Research, University of Sussex, Falmer, East Sussex, UK.
107. Laboratory of Neurological Diseases, Department of Neuroscience, IRCCS Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy.
108. Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of Bari 'Aldo Moro', Bari, Italy.
109. Unit of Neurodegenerative Diseases, Department of Clinical Research in Neurology, University of Bari 'Aldo Moro', at Pia Fondazione Cardinale G. Panico, Tricase, Lecce, Italy.
110. VIB, Vesalius Research Center, Laboratory of Neurobiology, Leuven, Belgium.
111. University Hospitals Leuven, Department of Neurology, Leuven, Belgium.
112. Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA.
113. Division of Genetics, Boston Children's Hospital, Boston, MA, USA.
114. Center for Basic Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA.
115. Department of Genetics, Harvard Medical School, Boston, MA, USA.
116. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.
117. Department of Pharmacology and Clinical Neurosience, Umeå University, Umeå, Sweden.
118. Centre SLA, CHRU de Tours, Tours, France.
119. Federation des Centres SLA Tours and Limoges, LITORALS, Tours, France.
120. Diamantina Institute, The University of Queensland, Translational Research Institute, Brisbane, Queensland, Australia.
121. Department of Epidemiology, Julius Center for Health Sciences and Primary Care,

University Medical Center Utrecht, Utrecht, The Netherlands.
122. Department of Medical and Molecular Genetics, King's College London, London, UK.
123. These authors contributed equally to this work.
124. These authors jointly directed this work.
125. To whom correspondence should be addressed.

Corresponding authors:
Ammar Al-Chalabi
Department of Basic and Clinical Neuroscience
King's College London
Maurice Wohl Clinical Neuroscience Institute
Coldharbour Lane
E-mail: ammar.al-chalabi@kcl.ac.uk

Jan H. Veldink
Department of Neurology
Brain Center Rudolf Magnus
University Medical Center Utrecht
Heidelberglaan 100
3584CX Utrecht
The Netherlands
Email: j.h.veldink@umcutrecht.nl

To elucidate the genetic architecture of amyotrophic lateral sclerosis (ALS) and find associated loci, we assembled a custom imputation reference panel from whole genome-sequenced ALS patients and matched controls (*N* = 1,861). Through imputation and mixed-model association analysis in 12,577 cases and 23,475 controls, combined with 2,579 cases and 2,767 controls in an independent replication cohort, we fine mapped a novel locus on chromosome 21 and identified *C21orf2* as an ALS risk gene. In addition, we identified *MOBP* and *SCFD1* as novel associated risk loci. We established evidence for ALS being a complex genetic trait with a polygenic architecture. Furthermore, we estimated the SNP-based heritability at 8.5%, with a distinct and important role for low frequency (1–10%) variants. This study motivates the interrogation of larger sample sizes with full genome coverage to identify rare causal variants that underpin ALS risk.

ALS is a fatal neurodegenerative disease that affects 1 in 400 people, death occurring within three to five years[1]. Twin-based studies estimate heritability to be around 65% and 5–10% of ALS patients have a positive family history[1,2]. Both are indicative of an important genetic component in ALS etiology. Following the initial discovery of the *C9orf72* locus in GWASs[3–5], the identification of the pathogenic hexanucleotide repeat expansion in this locus revolutionized the field of ALS genetics and biology[6,7]. The majority of ALS heritability, however, remains unexplained and only two additional risk loci have been identified robustly since[3,8].

To discover new genetic risk loci and elucidate the genetic architecture of ALS, we genotyped 7,763 new cases and 4,669 controls and additionally collected genotype data of published GWAS in ALS. In total, we analyzed 14,791 cases and 26,898 controls from 41 cohorts (**Supplementary Table 1**, **Supplementary Note**). We combined these cohorts based on genotyping platform and nationality to form 27 case-control strata. In total 12,577 cases and 23,475 controls passed quality control (Online methods, **Supplementary Tables 2–5**).

For imputation purposes we obtained high-coverage (~43.7X) whole genome sequencing data from 1,246 ALS patients and 615 controls from The Netherlands (Online methods, **Supplementary Fig. 1**). After quality control, we constructed a reference panel including 18,741,510 single nucleotide variants. Imputing this custom reference panel into Dutch ALS cases increased imputation accuracy of low-frequency variants (minor allele frequency, MAF 0.5–10%) considerably compared to commonly used reference panels: the 1000 Genomes

Project phase 1 (1000GP)[9] and Genome of The Netherlands (GoNL)[10] (**Fig. 1a**). The improvement was also observed when imputing into ALS cases from the UK (**Fig. 1b**). To benefit from the global diversity of haplotypes, the custom and 1000GP panels were combined, which further improved imputation. Given these results, we used the merged reference panel to impute all strata in our study.

In total we imputed 8,697,640 variants passing quality control in the 27 strata and separately tested these for association with ALS risk by logistic regression. Results were then included in an inverse-variance weighted fixed effects meta-analysis, which revealed 4 loci at genome-wide significance ($p < 5 \times 10^{-8}$) (**Fig. 2a**). The previously reported *C9orf72* (rs3849943)[3–5,8], *UNC13A (*rs12608932)[3,5] and *SARM1* (rs35714695)[8] loci all reached genome-wide significance, as did a novel association for a non-synonymous variant in *C21orf2* (rs75087725, $p = 8.7 \times 10^{-11}$, **Supplementary Tables 6–10**). Interestingly, this variant was present on only 10 haplotypes in the 1000GP reference panel (MAF = 1.3%), compared to 62 haplotypes in our custom reference panel (MAF = 1.7%). As a result, more strata passed quality control for this variant by passing the allele frequency threshold of 1% (**Supplementary Table 11**). This demonstrates the benefit of the merged reference panel with ALS-specific content, which improved imputation and resulted in a genome-wide significant association.

Linear mixed models (LMM) can improve power while controlling for sample structure[11], particularly in our study that included a large number of imperfectly balanced strata. Even though LMM for ascertained case-control data has a potential small loss of power[11], we judged the advantage of combining all strata while controlling the false positive rate, to be more important and therefore jointly analyzed all strata in a LMM to identify additional risk loci. There was no overall inflation of the linear mixed model's test statistic compared to the meta-analysis (**Supplementary Fig. 2**). We observed modest inflation in the QQ-plot ($\lambda_{GC} = 1.12$, $\lambda_{1000} = 1.01$, **Supplementary Fig. 3**). LD score regression yielded an intercept of 1.10 (standard error 7.8×10^{-3}). While the LD score regression intercept can indicate residual population stratification, which is fully corrected for in a LMM, the intercept can also reflect a distinct genetic architecture where most causal variants are rare, or a non-infinitesimal architecture[12]. The linear mixed model identified all four genome-wide significant associations from the meta-analysis. Furthermore, three additional loci that included the

MOBP gene on 3p22.1 (rs616147), *SCFD1* on 14q12 (rs10139154) and a long non-coding RNA on 8p23.2 (rs7813314) were associated at genome-wide significance (**Fig. 2b**, **Table 1**, **Supplementary Tables 12–14**). Interestingly, the SNPs in the *MOBP* locus have been reported in a GWAS on progressive supranuclear palsy (PSP)[13] and as a modifier for survival in frontotemporal dementia (FTD)[14]. The putative pleiotropic effect of variants within this locus suggests a shared neurodegenerative pathway between ALS, FTD and PSP. We also found rs74654358 at 12q14.2 in the *TBK1* gene approximating genome-wide significance (MAF = 4.9%, OR = 1.21 for A allele, $p = 6.6 \times 10^{-8}$). This gene was recently identified as an ALS risk gene through exome sequencing[15,16].

In the replication phase, we genotyped the newly discovered associated SNPs in nine independent replication cohorts, totaling 2,579 cases and 2,767 controls. In these cohorts we replicated the signals for the *C21orf2*, *MOBP* and *SCFD1* loci, with lower p-values in the combined analysis than the discovery phase (combined p-value = 3.08×10^{-10}, $p = 4.19 \times 10^{-10}$ and $p = 3.45 \times 10^{-8}$ for rs75087725, rs616147 and rs10139154 respectively, **Table 1**, **Supplementary Fig. 4**) [17]. The combined signal for rs7813314 was less significant due to an opposite effect between the discovery and replication phase, indicating non-replication. Although replication yielded similar effect estimates for rs10139154 compared to the discovery phase, this was not statistically significant (p = 0.09) in the replication phase alone. This reflects the limited sample size of our replication phase, which is inherent to the low prevalence of ALS and warrants even larger sample sizes to replicate this signal robustly.

There was no evidence for residual association within each locus after conditioning on the top SNP, indicating that all risk loci are independent signals. Apart from the *C9orf72, UNC13A* and *SARM1* loci, we found no evidence for associations previously described in smaller GWAS (**Supplementary Table 15**).

The associated low-frequency non-synonymous SNP in *C21orf2* suggested that this gene could directly be involved in ALS risk. Indeed, we found no evidence that linkage disequilibrium of sequenced variants beyond *C21orf2* explained the association within this locus (**Supplementary Fig. 5**). In addition, we investigated the burden of rare coding mutations in a set of whole genome sequenced cases (N = 2,562) and controls (N = 1,138). After quality control these variants were tested using a pooled association test for rare variants corrected for population structure (T5 and T1 for 5% and 1% allele frequency,

Supplementary Note). This revealed an excess of non-synonymous and loss-of-function mutations in *C21orf2* among ALS cases that persists after conditioning on rs75087725 (p_{T5} = 9.2×10^{-5}, p_{T1} = 0.01, **Supplementary Fig. 6**), which further supports that *C21orf2* contributes to ALS risk.

In an effort to fine-map the other loci to susceptibility genes, we searched for SNPs in these loci with *cis*-eQTL effects observed in brain and other tissues (**Supplementary Note**, **Supplementary Table 16**)[18]. There was overlap with previously identified brain *cis*-eQTLs for five regions (**Supplementary Fig. 7**, **Supplementary Table 17, Supplementary Data Set 1**). Interestingly, within the *C9orf72* locus we found that proxies of rs3849943 (LD r^2 = 0.21 - 0.56) had a brain *cis*-eQTL effect on *C9orf72* only (minimal p = 5.27×10^{-7}), which harbors the hexanucleotide repeat expansion that drives this GWAS signal. Additionally, we found that rs12608932 and its proxies within the *UNC13A* locus had exon-level *cis*-eQTL effect on *KCNN1* in frontal cortex (p = 1.15×10^{-3})[19]. Another overlap was observed in the *SARM1* locus where rs35714695 and its proxies had the strongest exon-level *cis*-eQTL effect on *POLDIP2* in multiple brain tissues (p = 2.32×10^{-3}). Within the *SCFD1* locus rs10139154 and proxies had a *cis*-eQTL effect on *SCFD1* in cerebellar tissue (p = 7.71×10^{-4}). For the *MOBP* locus, rs1768208 and proxies had a *cis*-eQTL effect on *RPSA* (p = 7.71×10^{-4}).

To describe the genetic architecture of ALS, we calculated polygenic scores that can be used to predict phenotypes for traits with a polygenic architecture[20]. We calculated the SNP effects using a linear mixed model in 18 of the 27 strata and subsequently assessed their predictive ability in the other 9 independent strata. This revealed that a significant, albeit modest, proportion of the phenotypic variance could be explained by all SNPs (Nagelkerke r^2 = 0.44%, r^2 = 0.15% on the liability scale, p = 2.7×10^{-10}, **Supplementary Fig. 8**). This finding adds to the existing evidence that ALS is a complex genetic trait with a polygenic architecture. To further quantify the contribution of common SNPs to ALS risk, we estimated the SNP-based heritability using three approaches, all assuming a population baseline risk of 0.25%[21]. GCTA-REML estimated the SNP-based heritability at 8.5% (SE 0.5%). Haseman-Elston regression yielded a very similar 7.9% and LD score regression estimated the SNP-based heritability at 8.2% (SE 0.5%). The heritability estimates per chromosome were strongly correlated with chromosome length (p = 4.9×10^{-4}, r^2 = 0.46, **Fig. 3a**), which again is indicative of the polygenic architecture of ALS.

We found that the genome-wide significant loci only explained 0.2% of the heritability and thus the bulk of the heritability (8.3%, SE 0.3%) was captured in SNPs below genome-wide significance. This implies that many genetic risk variants have yet to be discovered. Understanding where these unidentified risk variants remain across the allele frequency spectrum will inform designing future studies to identify these variants. We, therefore, estimated heritability partitioned by minor allele frequency. Furthermore, we contrasted this to common polygenic traits studied in GWASs such as schizophrenia. We observed a clear trend that indicated that most variance is explained by low-frequency SNPs (**Fig. 3b**). Exclusion of the *C9orf72* locus, which harbors the rare pathogenic repeat expansion, and the other genome-wide significant loci did not affect this trend (**Supplementary fig. 9**). This architecture is different from that expected for common polygenic traits and reflects a polygenic rare-variant architecture observed in simulations[22].

To gain better insight into the biological pathways that explain the associated loci found in this study we looked for enriched pathways using DEPICT[23]. This revealed SNAP receptor (SNARE) activity as the only enriched category (FDR < 0.05, **Supplementary Fig. 10**). SNARE complexes play a central role in neurotransmitter release and synaptic function[24], which are both perturbed in ALS[25].

Although the biological role of C21orf2, a conserved leucine-rich repeat protein, remains poorly characterized, it is part of the ciliome and is required for the formation and/or maintenance of primary cilia[26]. Defects in primary cilia are associated with various neurological disorders and cilia numbers are decreased in G93A *SOD1* mice, a well-characterized ALS model[27]. C21orf2 has also been localized to mitochondria in immune cells[28] and is part of the interactome of the protein product of *NEK1*, which has previously been associated with ALS[15]. Both proteins appear to be involved in DNA repair mechanisms[29]. Although future studies are needed to dissect the function of *C21orf2* in ALS pathophysiology it is tempting to speculate that defects in *C21orf2* lead to primary cilium and/or mitochondrial dysfunction or inefficient DNA repair and thereby adult onset disease. The other associated loci will require more extensive studies to fine-map causal variants. The *SARM1* gene has been suggested as a susceptibility gene for ALS, mainly because of its role in Wallerian degeneration and interaction with *UNC13A*[8,30]. Although these are indeed interesting observations, the brain *cis*-eQTL effect on *POLDIP2* suggests that *POLDIP2* and not *SARM1* could in fact be the causal gene within this locus. Similarly, *KCNN1*, which

encodes a neuronal potassium channel involved in neuronal excitability, could be the causal gene either through a direct eQTL effect or rare variants in LD with the associated SNP in *UNC13A*.

In conclusion, we identified a key role for rare variation in ALS and discovered SNPs in novel complex loci. Our study therefore informs future study design in ALS genetics: the combination of larger sample sizes, full genome coverage and targeted genome editing experiments, leveraged together to fine map novel loci, identify rare causal variants and thereby elucidate the biology of ALS.

ACCESSION CODES

NIH Genome-Wide Association Studies of Amyotrophic Lateral Sclerosis (phs000101.v3.p1), Genome-Wide Association Study of Amyotrophic Lateral Sclerosis in Finland (phs000344.v1.p1), CIDR: Genome Wide Association Study in Familial Parkinson Disease (PD) (phs000126.v1.p1), Genome-Wide Association Study of Parkinson Disease: Genes and Environment (phs000196.v1.p1)

DATA ACCESS

The GWAS summary statistics and sequenced variants are publicly available through the Project MinE data browser: http://databrowser.projectmine.com

AUTHOR INFORMATION

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to A.A-C or J.H.V. (ammar.al-chalabi@kcl.ac.uk or j.h.veldink@umcutrecht.nl).

ACKNOWLEDGMENTS

The work of the contributing groups was supported by various grants from governmental and charitable bodies. Details are provided in the **Supplementary Notes**.

AUTHOR CONTRIBUTIONS

A.V., N.T., K.L., B.R, K.V, M.R-G, B.K, J.Z, L.L., L.D.G., S.M., F.S, V.M., M.d.C, S.Pinto, J.M, R.R-G, M.P., S.Chandran., S.Colville, R.S, K.E.M., P.J.S., J.H., R.W.O., A.Pittman., K.S., P.F., A.M., S.T., S.Petri, S.Abdulla., C.D., M.S., T.Meyer., R.A.O., K.A.S., M.W-P., C.L-H., V.M.V.D, J.Q.T, L.E, L.McC., A.N.B., Y.P., T.Meitinger, P.L., M.B-R., C.R.A., C.Maurel, G.B., B.L., A.B., C.A.M.P., S.S-D., N.W.W., L.T., W.L., A.F., M.R., S.C., M.M.N., P.A., C.Tzourio., J-F.D., A.G.U., F.R., K.E., A.H., C.Curtis, H.M.B., A.J.v.d.K., M.d.V., A.G., M.W., C.E.S., B.N.S., O.P., C.Cereda, R.D.B., G.P.C., S.D'A., C.B., G.S, L.M., V.P., C.G., C.Tiloca, A.R., A.Calvo., C.Moglia, M.B., S.Arcuti., R.C., C.Z., C.L., S.Penco, N.R., A.Padovani., M.F., B.M., R.J.S., PARALS registry, SLALOM group, SLAP registry, FALS sequencing consortium, SLAGEN consortium, NNIPPS study group, I.B., G.A.N., D.B.R., R.P., M.C.K., J.G.., O.W.W., T.R., T.A.P., B.S., I.K., C.A.H., P.N.L., F.C., A.Chìo., E.B., E.P., R.T., G.L., J.P., A.C.L., J.H.W., W.R., P.V.D., L.F., T.P., R.H.B., J.D.G, J.E.L., O.H., P.M.A., P.C., P.V., V.S., M.A.v.E., A.A.-C, L.H.v.d.B. and J.H.V. were involved in

phenotyping, sample collection and management. W.v.R, A.S., A.M.D., R.L.M., F.P.D., R.A.A.v.d.S., P.T.C.v.D., G.H.P.T., M.K., A.M.B., W.S., A.R.J., K.P.K., I.F., A.V., N.T., R.D.S., W.J.B., A.V., K.V., M.R.-G., B.K, L.L., S.Abdulla., K.S., E.P., F.P.D., J.M., C.Curtis, G.B., A.A.-C and J.H.V. prepared DNA and performed SNP-array hybridizations. W.v.R., S.L.P., K.K., K.L., A.M.D., P.T.C.v.D., G.H.P.T., K.R.v.E., P.I.W.d.B and J.H.V. were involved in the next-generations sequencing analyses. W.v.R., K.R.v.E., A.M., P.I.W.d.B., A.A. and J.H.V. performed the imputation. W.v.R., A.S., F.P.D., R.L.M., S.L.P., S.d.J., I.F., N.T., W.S., A.J., K.P.K., K.R.v.E., K.S., H.M.B., P.I.W.d.B., M.A.v.E., C.L., G.B., A.A.-C., L.H.v.d.B and J.H.V. performed GWAS analyses. W.v.R., A.M.D., R.A.A.v.d.S., R.L.M., C.A., M.K., A.M.B., R.D.S., E.P.M., J.A.F., C.Tzourio, H.H., K.Z., P.C., P.V. and J.H.V. performed the replication analyses. W.v.R., A.S., R.L.M., M.R.R., J.Y., N.R.W., P.M.V., C.L., A.A.-C and J.H.V. performed polygenic risk scoring and heritability analyses. S.d.J., U.V., L.F., T.P., W.v.R., O.H., G.B., R.J.P. and J.H.V. performed biological pathway analyses. U.V., L.F., W.v.R. and J.H.V. performed eQTL analyses. W.v.R., A.S., A.A.-C., L.H.v.d.B. and J.H.V., prepared the manuscript with contributions from all authors. A.A.-C., L.H.v.d.B. and J.H.V. directed the study.

REFERENCES FOR MAIN TEXT

1. Hardiman, O., van den Berg, L. H. & Kiernan, M. C. Clinical diagnosis and management of amyotrophic lateral sclerosis. *Nat Rev Neurol* **7,** 639–649 (2011).

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

3. van Es, M. A. *et al*. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* **41,** 1083–1087 (2009).

4. Laaksovirta, H. *et al*. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol.* **9,** 978–985 (2010).

5. Shatunov, A. *et al*. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol.* **9,** 986–994 (2010).

6. DeJesus-Hernandez, M. *et al*. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72,** 245–256 (2011).

7. Renton, A. E. *et al*. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72,** 257–268 (2011).
8. Fogh, I. *et al*. A genome-wide association meta-analysis identifies a novel locus at 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **23,** 2220–2231 (2014).
9. 1000 Genomes Project Consortium *et al*. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
10. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46,** 818–825 (2014).
11. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46,** 100–106 (2014).
12. Bulik-Sullivan, B. K. *et al*. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
13. Höglinger, G. U. *et al*. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* (2011).
14. Irwin, D. J. *et al*. Myelin oligodendrocyte basic protein and prognosis in behavioral-variant frontotemporal dementia. *Neurology* **83,** 502–509 (2014).
15. Cirulli, E. T. *et al*. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* **347,** 1436–1441 (2015).
16. Freischmidt, A. *et al*. Haploinsufficiency of TBK1 causes familial ALS and fronto-temporal dementia. *Nat. Neurosci.* **18,** 631–636 (2015).
17. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38,** 209–213 (2006).
18. Nicolae, D. L. et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, (2010).
19. Ramasamy, A. *et al*. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17,** 1418–1428 (2014)
20. Wray, N. R. *et al*. Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14,** 507–515 (2013).
21. Johnston, C. A. *et al*. Amyotrophic lateral sclerosis in an urban setting: a population based study of inner city London. *J. Neurol.* **253,** 1642–1643 (2006).

22. Lee, S. H. *et al*. Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. *Nat. Genet*. **44,** 247–250 (2012).
23. Pers, T. H. *et al*. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun*. **6,** 5890–20 (2015).
24. Ramakrishnan, N. A., Drescher, M. J. & Drescher, D. G. The SNARE complex in neuronal and sensory cells. *Mol. Cell. Neurosci*. **50,** 58–69 (2012).
25. Ferraiuolo, L., Kirby, J., Grierson, A. J., Sendtner, M. & Shaw, P. J. Molecular pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat. Rev. Neurol*. **7,** 616–630 (2011).
26. Lai, C. K. *et al*. Functional characterization of putative cilia genes by high-content analysis. *Mol. Biol. Cell*. **22,** 1104–1119 (2011).
27. Ma, X., Peterson, R. & Turnbull, J. Adenylyl cyclase type 3, a marker of primary cilia, is reduced in primary cell culture and in lumbar spinal cord in situ in G93A SOD1 mice. *BMC. Neurosci*. **12,** 71 (2011).
28. Krohn, K., Ovod, V., Vilja, P., Heino, M. & Scott, H. Immunochemical characterization of a novel mitochondrially located protein encoded by a nuclear gene within the DFNB8/10 critical region on 21q22. 3. *Biochem. Biophys. Res. Commun*. (1997).
29. Fang, X. *et al*. The NEK1 interactor, C21ORF2, is required for efficient DNA damage repair. *Acta Biochim. Biophys. Sin*. **47,** 834–841 (2015)
30. Vériène, J., Fossouo, L. & Parker, J. A. Neurodegeneration in C. elegans models of ALS requires TIR-1/Sarm1 immune pathway activation in neurons. *Nat. Commun*. **6,** 7319 (2015).

FIGURE LEGENDS

Figure 1. Imputation accuracy comparison. The aggregate r^2 value between imputed and sequenced genotypes on chromosome 20 using different reference panels for imputation. Allele frequencies are calculated from the Dutch samples included in the Genome of the Netherlands cohort. The highest imputation accuracy was achieved when imputing from the merged custom and 1000GP panels. This difference is most pronounced for low frequency (0.5–10%) alleles in both ALS cases from The Netherlands **(a)** and United Kingdom **(b)**.

Figure 2. Meta-analysis and linear mixed model associations. (**a**) Manhattan plot for meta-analysis results. This yielded four genome-wide significant associations highlighted with names indicating the closest gene. The associated SNP in *C21orf2* is a non-synonymous

variant not found in previous GWAS. (**b)** Manhattan plot for linear mixed model results. This association analysis yielded three additional loci reaching genome-wide significance (*MOBP, LOC101927815 and SCFD1)*. SNPs in the previously identified ALS risk gene *TBK1* approached genome-wide significance ($p = 6.6 \times 10^{-8}$). Since the *C21orf2* SNP was removed from a Swedish stratum because of a MAF < 1%, this SNP was tested separately, but is presented here together with all other SNPs with a MAF > 1% in every stratum. Here, *LOC101927815* is colored grey because the association for this locus could not be replicated.

Figure 3. Partitioned heritability. (a) The heritability estimates per chromosome were strongly correlated with chromosome length ($p = 4.9 \times 10^{-4}$). **(b)** For ALS there was a clear trend where more heritability was explained within the lower allele frequency bins. This effect was still observed when, for a fair comparison between ALS and a previous study partitioning heritability for schizophrenia (SCZ) using identical methods[22], SNPs present in HapMap3 (HM3) were included. The pattern for ALS resembles that observed in a rare variant model simulation performed in this study. Error bars reflect standard errors.

TABLES

Table 1. Discovery and replication of novel genome-wide significant loci.

	Discovery					Replication				Combined	
SNP	MAF_{cases}	$MAF_{controls}$	*OR*	P_{meta}	P_{LMM}	MAF_{cases}	$MAF_{controls}$	*OR*	*P*	$P_{combined}$	I^2
rs75087725	0.02	0.01	1.45	8.65×10^{-11}	2.65×10^{-9}	0.02	0.01	1.65	3.89×10^{-3}	3.08×10^{-10}	0.00*
rs616147	0.30	0.28	1.10	4.14×10^{-5}	1.43×10^{-8}	0.31	0.28	1.13	2.35×10^{-3}	4.19×10^{-10}	0.00*
rs10139154	0.34	0.31	1.09	1.92×10^{-5}	4.95×10^{-8}	0.33	0.31	1.06	9.55×10^{-2}	3.45×10^{-8}	0.05*
rs7813314	0.09	0.10	0.87	7.46×10^{-7}	3.14×10^{-8}	0.12	0.10	1.17	7.75×10^{-3}	1.05×10^{-5}	0.80**

Table 1. Discovery and replication of novel genome-wide significant loci. Genome-wide significant loci from the discovery phase including 12,557 cases and 23,475 controls were directly genotyped and tested for association in the replication phase including 2,579 cases and 2,767 controls. The three top associated SNPs in the *MOBP* (rs616147), *SCFD1 (*rs10139154) and *C21orf2 (*rs75087725) loci replicated with associations in identical directions as in the discovery phase and an association in the combined analysis that exceeded the discovery phase. * Cochrane's Q test: $p > 0.1$, ** Cochrane's Q test: $p = 4.0 \times 10^{-6}$, Chr = chromosome; SNP = single nucleotide polymorphism, MAF = minor allele frequency, OR = odds ratio, P_{meta} = meta-analysis p-value, P_{LMM} = linear mixed model p-value, $P_{combined}$ = meta-analysis of discovery linear mixed model and associations from replication phase.

Figure 1

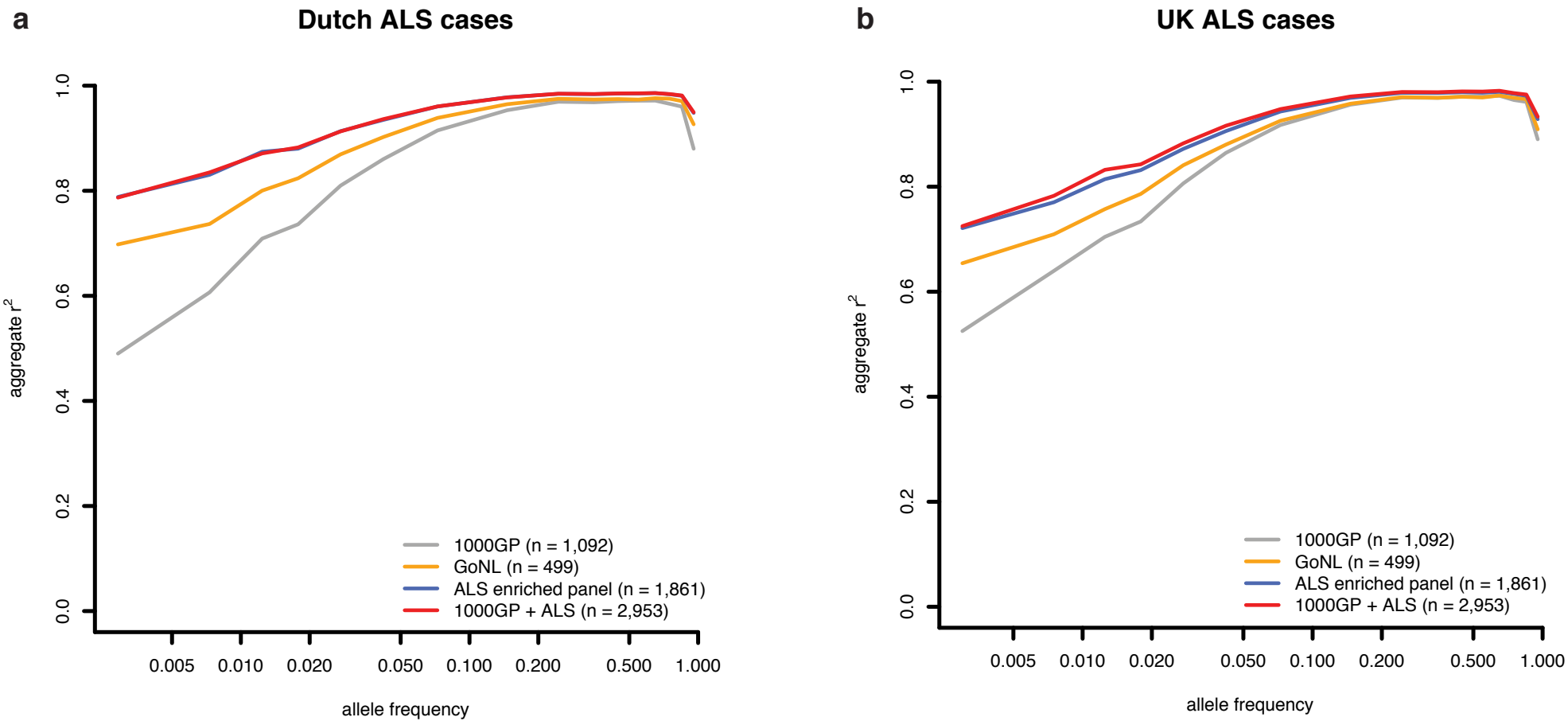

a
Dutch ALS cases
b
UK ALS cases
aggregate r2
allele frequency
0.0
0.2
0.4
0.6
0.8
1.0
0.005
0.010
0.020
0.050
0.100
0.200
0.500
1.000
1000GP (n = 1,092)
GoNL (n = 499)
ALS enriched panel (n = 1,861)
1000GP + ALS (n = 2,953)

Figure 2

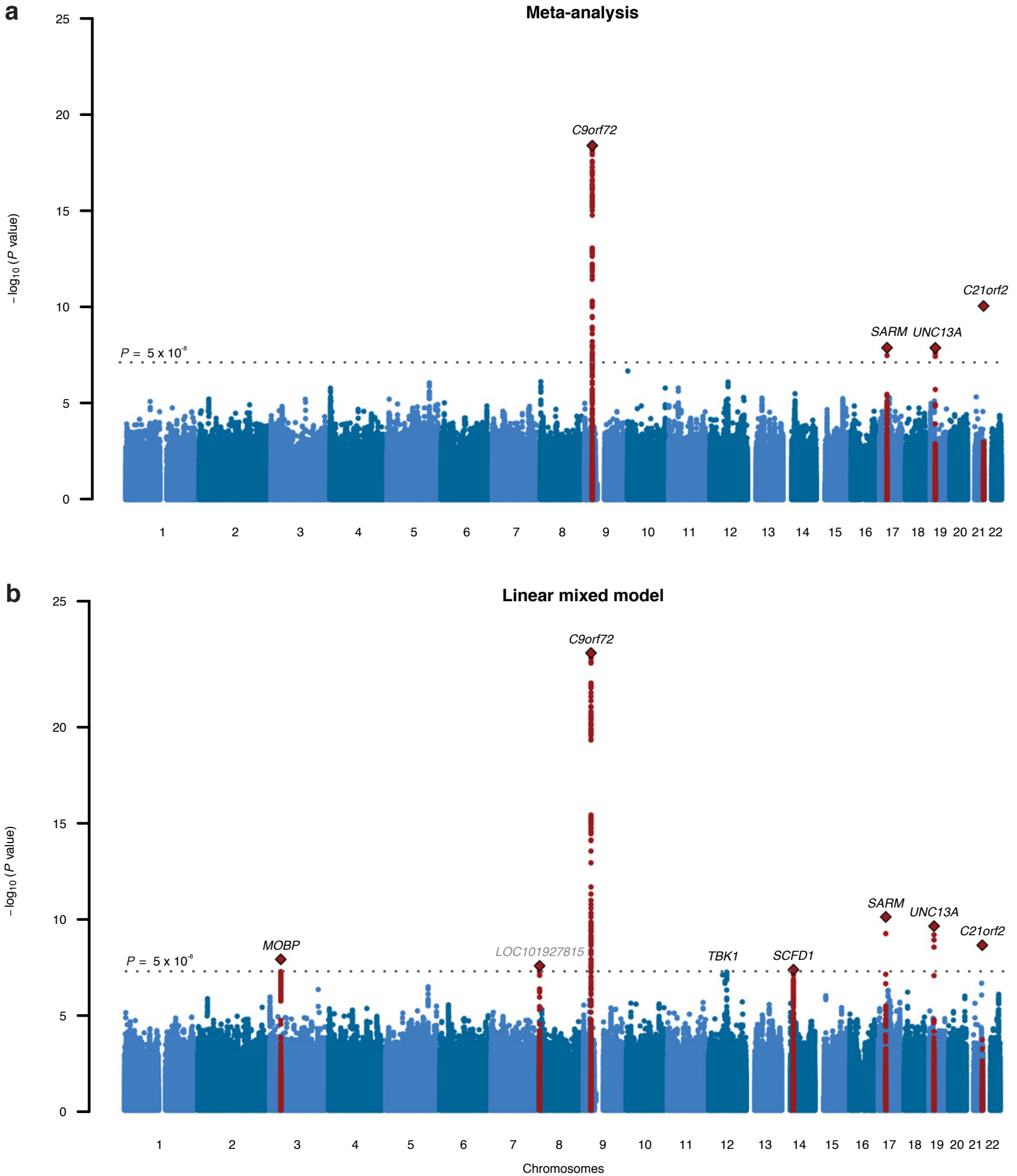
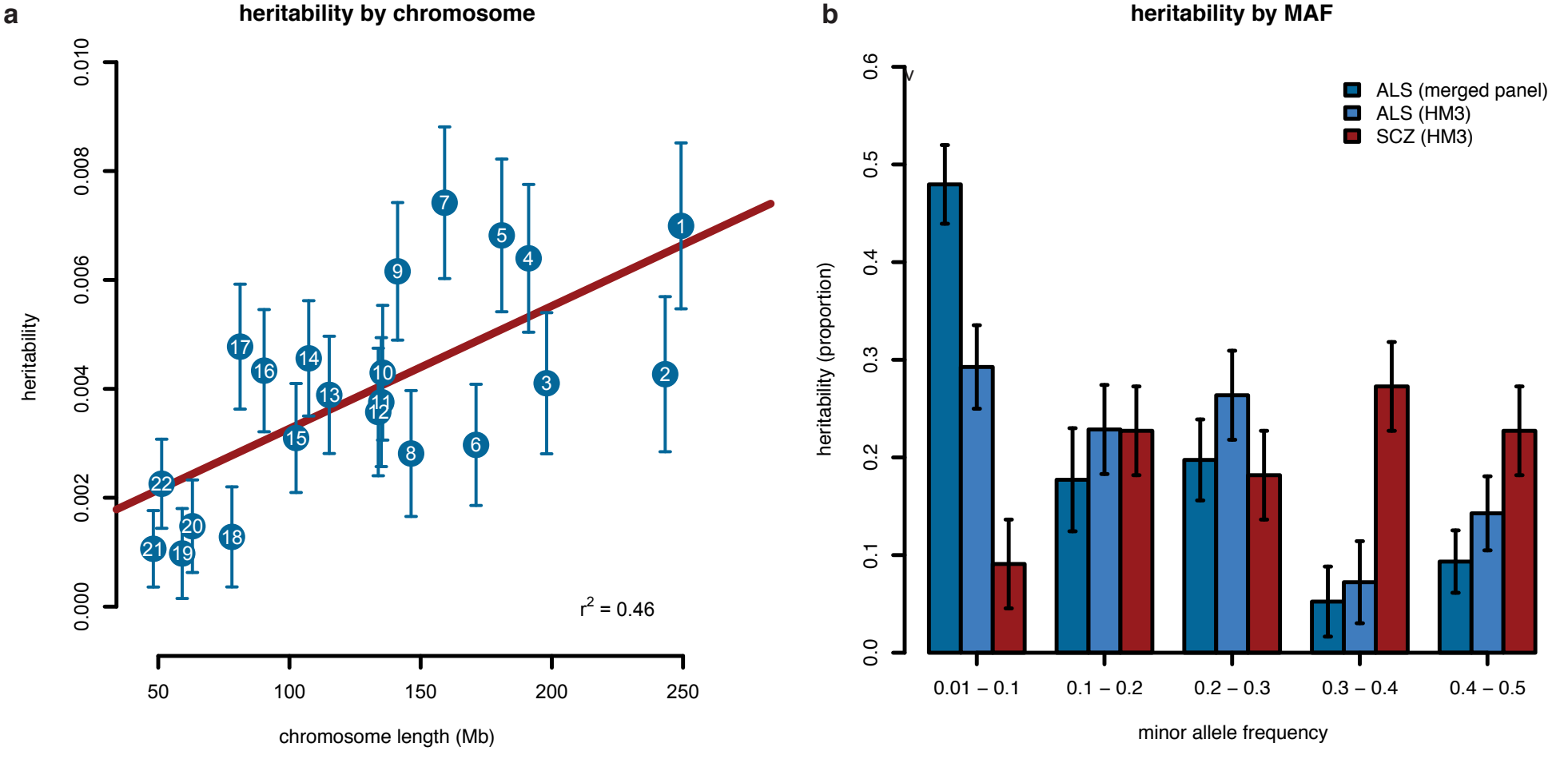

a
Meta-analysis
C9orf72
C21orf2
SARM
UNC13A
P = 5 x 10-8
−log10 (P value)
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
b
Linear mixed model
C9orf72
MOBP
LOC101927815
TBK1
SCFD1
SARM
UNC13A
C21orf2
P = 5 x 10-8
−log10 (P value)
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosomes

Figure 3


a
heritability by chromosome
heritability
0.000
0.002
0.004
0.006
0.008
0.010
50
100
150
200
250
chromosome length (Mb)
r2 = 0.46
b
heritability by MAF
heritability (proportion)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
ALS (merged panel)
ALS (HM3)
SCZ (HM3)
0.01 – 0.1
0.1 – 0.2
0.2 – 0.3
0.3 – 0.4
0.4 – 0.5
minor allele frequency

ONLINE METHODS

Software packages used, their version, web source, and references are described in the **Supplementary Table 18**.

GWAS discovery phase and quality control. Details on the acquired genotype data from previously published GWAS are described in **Supplementary Table 1**. Methods for case and control ascertainment for each cohort are described in the **Supplementary Note**. All cases and controls gave written informed consent and the relevant institutional review boards approved this study. To obtain genotype data for newly genotyped individuals, genomic DNA was hybridized to the Illumina OmniExpress array according to manufacturer's protocol. Subsequent quality control included:

1) Removing low quality SNPs and individuals from each cohort.
2) Combining unbalanced cohorts based on nationality and genotyping platform to form case-control strata.
3) Removing low quality SNPs, related individuals and population outliers per stratum.
4) Calculate genomic inflation factors per stratum.

More details are described in the **Supplementary Note** and **Supplementary Fig. 11**. The number of SNPs and individuals failing each QC step per cohort and stratum are displayed in **Supplementary Tables 2–5**.

Whole genome sequencing (custom reference panel). Individuals were whole genome sequenced on the Illumina HiSeq 2500 platform using PCR free library preparation and 100bp paired-end sequencing yielding a minimum 35X coverage. Reads were aligned to the hg19 human genome build and after variant calling (Isaac variant caller) additional SNV and sample quality control was performed (**Supplementary Note** and **Supplementary Fig. 12**). Individuals in our custom reference panel were also included in the GWAS in strata sNL2, sNL3 and sNL4.

Merging reference panels. All high quality calls in the custom reference panel were phased using SHAPEIT2 software. After checking strand and allele inconsistencies, both the 1000 Genomes Project (1000GP) reference panel (release 05-21-2011)[31] and custom reference panel were imputed up to the union of their variants as described previously[32]. Those variants with inconsistent allele frequencies between the two panels were removed.

Imputation accuracy performance. To assess the imputation accuracy between different reference panels, 109 unrelated ALS cases of Dutch ancestry sequenced by Complete Genomics and 67 ALS cases from the UK sequenced by Illumina were selected as a test panel. All variants not present on the Illumina Omni1M array were masked and the SNVs on chromosome 20 were subsequently imputed back using four different reference panels (1000GP, GoNL, custom panel and merged panel). Concordance between the imputed alleles and sequenced alleles was assessed within each allele frequency bin where allele frequencies are calculated from the Dutch samples included in the Genome of the Netherlands cohort.

GWAS imputation. Pre-phasing was performed per stratum using SHAPEIT2 with the 1000GP phase 1 (release 05-21-2011) haplotypes[31] as a reference panel. Subsequently, strata were imputed up to the merged reference panel in 5 megabase chunks using IMPUTE2. Imputed variants with a MAF < 1% or INFO score < 0.3 were excluded from further analysis. Variants with allele frequency differences between strata, defined as deviating > 10SD from the normalized mean allele frequency difference between those strata and an absolute difference > 5%, were excluded, since they are likely to represent sequencing or genotyping artifacts. Imputation concordance scores for cases and controls were compared to assess biases in imputation accuracy (**Supplementary Table 19**).

Meta-analysis. Logistic regression was performed on imputed genotype dosages under an additive model using SNPTEST software. Based on scree plots, 1 to 4 principal components were included per stratum. These results were then combined in an inverse-variance weighted fixed effect meta-analysis using METAL. No marked heterogeneity across strata was observed as the Cochrane's Q test statistics did not deviate from the null-distribution (λ = 0.96). Therefore, no SNPs were removed due to excessive heterogeneity. The genomic inflation factor was calculated and the quantile-quantile plot is provided in **Supplementary Fig. 3a**.

Linear mixed model. All strata were combined including SNPs that passed quality control in every stratum. Subsequently the genetic relationship matrices (GRM) were calculated per chromosome including all SNPs using the Genome-Wide Complex Trait Analysis (GCTA) software package. Each SNP was then tested in a linear mixed model including a GRM

composed of all chromosomes excluding the target chromosome (leave one chromosome out, LOCO). The genomic inflation factor was calculated and the quantile-quantile plot is provided as **Supplementary Fig. 3b**.

Replication. For the replication phase independent ALS cases and controls from Australia, Belgium, France, Germany, Ireland, Italy, The Netherlands and Turkey that were not used in the discovery phase were included. A pre-designed TaqMan genotyping assay was used to replicate rs75087725 and rs616147. Sanger sequencing was performed to replicate rs10139154 and rs7813314 (**Supplementary Note** and **Supplementary Table 20**). All genotypes were tested in a logistic regression per country and subsequently meta-analyzed.

Rare variant analysis in *C21orf2*. The burden of non-synonymous rare variants in *C21orf2* was assessed in whole genome sequencing data obtained from ALS cases and controls from The Netherlands, Belgium, Ireland, United Kingdom and the United States. After quality control the burden of non-synonymous and loss-of-function mutations in *C21orf2* were tested for association per country and subsequently meta-analyzed. More details are provided in the **Supplementary Note** and **Supplementary Fig. 13**.

Polygenic risk scores. To assess the predictive accuracy of polygenic risk scores in an independent dataset SNP weights were assigned based on the linear mixed model (GCTA-LOCO) analysis in 18/27 strata. SNPs in high LD ($r^2 > 0.5$) within a 250 kb window were clumped. Subsequently, polygenic risk scores for cases and controls in the 9 independent strata were calculated based on their genotype dosages using PLINK v1.9. To obtain the Nagelkerke R^2 and corresponding p-values these scores were then regressed on their true phenotype in a logistic regression where (based on scree plots) the first three PCs, sex and stratum were included as covariates.

SNP-based heritability estimates. *GCTA-REML*. GRMs were calculated using GCTA software including genotype dosages passing quality control in all strata. Based on the diagonal of the GRM individuals representing subpopulations that contain an abundance of rare alleles (diagonal values mean +/- 2SD) were removed (**Supplementary Fig. 14a**). Pairs where relatedness (off-diagonal) exceeded 0.05 were removed as well (**Supplementary Fig. 14b**). The eigenvectors for the first 10 PCs were included as fixed effects to account for more subtle population structure. The prevalence of ALS was defined as the life-time morbid risk

for ALS (i.e. 1/400)[19]. To estimate the SNP-based heritability for all non-genome-wide significant SNPs, genotypes for the SNPs reaching genome-wide significance were modeled as fixed effect. The variance explained by the GRM therefore reflects the SNP-based heritability of all non-genome-wide significant SNPs. SNP-based heritability partitioned by chromosome or MAF was calculated by including multiple GRMs, calculated on SNPs from each chromosome or within the respective frequency bin, in one model.

Haseman-Elston regression. The Phenotype correlation - Genotype correlation (PCGC) regression software package was used to calculate heritability based on the Haseman-Elston regression including the eigenvectors for the first 10 PCs as covariates. The prevalence was again defined as the life-time morbid risk (1/400).

LD score regression. Summary statistics from GCTA-LOCO and LD scores calculated from European individuals in 1000GP were used for LD score regression. Strongly associated SNPs ($p < 5 \times 10^{-8}$) and variants not in HapMap3 were excluded. Considering adequate correction for population structure and distant relatedness in the linear mixed model, the intercept was constrained to 1.0[12].

Biological pathway analysis (DEPICT). Functional interpretation of associated GWAS loci was carried out using DEPICT, using locus definition based on 1000GP phase 1 data. This method prioritizes genes in the affected loci, predicts involved pathways, biological processes and tissues, using gene co-regulation data from 77,840 expression arrays. Three separate analyses were performed for GWAS loci reaching $p = 10^{-4}$, $p = 10^{-5}$ or $p = 10^{-6}$. One thousand permutations were used for adjusting the nominal enrichment p-values for biases and additionally 200 permutations were used for FDR calculation.

REFERENCES FOR METHODS

31. Delaneau, O., Marchini, J. & 1000 Genomes Project Consortium. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun*. **5,** 3934 (2014).

32. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1,** 457-70 (2011).