



Edinburgh Research Explorer

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

Citation for published version:

Van Rheenen, W, Shatunov, A, Dekker, AM, Mclaughlin, RL, Diekstra, FP, Pulit, SL, Van Der Spek, RAA, Vösa, U, De Jong, S, Robinson, MR, Yang, J, Fogh, I, Van Doormaal, PT, Tazelaar, GHP, Koppers, M, Blokhuis, AM, Sproviero, W, Jones, AR, Kenna, KP, Van Eijk, KR, Harschnitz, O, Schellevis, RD, Brands, WJ, Medic, J, Menelaou, A, Vajda, A, Ticozzi, N, Lin, K, Rogelj, B, Vrabec, K, Ravnik-glava, M, Koritnik, B, Zidar, J, Leonardis, L, Grošelj, LD, Millecamps, S, Salachas, F, Meininger, V, De Carvalho, M, Pinto, S, Mora, JS, Rojas-garcía, R, Polak, M, Chandran, S, Colville, S, Swingler, R, Morrison, KE, Shaw, PJ, Hardy, J, Orrell, RW, Pittman, A, Sidle, K, Fratta, P, Malaspina, A, Topp, S, Petri, S, Abdulla, S, Drepper, C, Sendtner, M, Meyer, T, Ophoff, RA, Staats, KA, Wiedau-pazos, M, Lomen-hoerth, C, Van Deerlin, VM, Trojanowski, JQ, Elman, L, Mccluskey, L, Basak, AN, Tunca, C, Hamzeiy, H, Parman, Y, Meitinger, T, Lichtner, P, Radivojkov-bлагоjevic, M, Andres, CR, Maurel, C, Bensimon, G, Landwehrmeyer, B, Brice, A, Payan, CAM, Saker-delye, S, Dürr, A, Wood, NW, Tittmann, L, Lieb, W, Franke, A, Rietschel, M, Cichon, S, Nöthen, MM, Amouyel, P, Tzourio, C, Dartigues, J, Uitterlinden, AG, Rivadeneira, F, Estrada, K, Hofman, A, Curtis, C, Blauw, HM, Van Der Kooij, AJ, De Visser, M, Goris, A, Weber, M, Shaw, CE, Smith, BN, Pansarasa, O, Cereda, C, Del Bo, R, Comi, GP, D'alfonso, S, Bertolin, C, Sorarù, G, Mazzini, L, Pensato, V, Gellera, C, Tiloca, C, Ratti, A, Calvo, A, Moglia, C, Brunetti, M, Arcuti, S, Capozzo, R, Zecca, C, Lunetta, C, Penco, S, Riva, N, Padovani, A, Filosto, M, Muller, B, Stuit, RJ, Blair, I, Zhang, K, Mccann, EP, Fifita, JA, Nicholson, GA, Rowe, DB, Pamphlett, R, Kiernan, MC, Grosskreutz, J, Witte, OW, Ringer, T, Prell, T, Stübendorff, B, Kurth, I, Hübner, CA, Leigh, PN, Casale, F, Chio, A, Beghi, E, Pupillo, E, Tortelli, R, Logroscino, G, Powell, J, Ludolph, AC, Weishaupt, JH, Robberecht, W, Van Damme, P, Franke, L, Pers, TH, Brown, RH, Glass, JD, Landers, JE, Hardiman, O, Andersen, PM, Corcia, P, Vourc'h, P, Silani, V, Wray, NR, Visscher, PM, De Bakker, PIW, Van Es, MA, Pasterkamp, RJ, Lewis, CM, Breen, G, Al-chalabi, A, Van Den Berg, LH & Veldink, JH 2016, 'Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis', *Nature Genetics*. <https://doi.org/10.1038/ng.3622>

Digital Object Identifier (DOI):

[10.1038/ng.3622](https://doi.org/10.1038/ng.3622)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature Genetics

Publisher Rights Statement:

This is author's peer-reviewed manuscript as accepted for publication

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.



1 Genome-wide association analyses identify new risk variants 2 and the genetic architecture of amyotrophic lateral sclerosis

3
4 Wouter van Rheenen^{1,125}, Aleksey Shatunov^{2,125}, Annelot M. Dekker¹, Russell L. McLaughlin³,
5 Frank P. Diekstra¹, Sara L. Pulit⁴, Rick A.A. van der Spek¹, Urmo Vösa⁵, Simone de Jong^{6,7},
6 Matthew R. Robinson⁸, Jian Yang⁸, Isabella Fogh^{2,9}, Perry T.C. van Doormaal¹, Gijs H.P.
7 Tazelaar¹, Max Koppers^{1,10}, Anna M. Blokhuis^{1,10}, William Sproviero², Ashley R. Jones²,
8 Kevin P. Kenna¹¹, Kristel R. van Eijk¹, Oliver Harschnitz^{1,10}, Raymond D. Schellevis¹,
9 William J. Brands¹, Jelena Medic¹, Androniki Menelaou⁴, Alice Vajda^{12,13}, Nicola Ticozzi^{9,14},
10 Kuang Lin², Boris Rogelj^{15,16}, Katarina Vrabec¹⁷, Metka Ravnik-Glavac^{17,18}, Blaž Koritnik¹⁹,
11 Janez Zidar¹⁹, Lea Leonardis¹⁹, Leja Dolenc Grošelj¹⁹, Stéphanie Millecamps²⁰, François
12 Salachas^{21,20,22}, Vincent Meininger^{23,24}, Mamede de Carvalho^{25,26}, Susana Pinto^{25,26}, Jesus S.
13 Mora²⁷, Ricardo Rojas-García^{28,29}, Meraida Polak^{30,31}, Siddharthan Chandran^{32,33}, Shuna
14 Colville³², Robert Swingler³², Karen E. Morrison^{34,35}, Pamela J. Shaw³⁶, John Hardy³⁷, Richard
15 W. Orrell³⁸, Alan Pittman^{37,39}, Katie Sidle³⁸, Pietro Fratta⁴⁰, Andrea Malaspina^{41,42}, Simon
16 Topp², Susanne Petri⁴³, Susanne Abdulla⁴⁴, Carsten Drepper⁴⁵, Michael Sendtner⁴⁵, Thomas
17 Meyer⁴⁶, Roel A. Ophoff^{47,48,49}, Kim A. Staats⁴⁹, Martina Wiedau-Pazos⁵⁰, Catherine Lomen-
18 Hoerth⁵¹, Vivianna M. Van Deerlin⁵², John Q. Trojanowski⁵², Lauren Elman⁵³, Leo
19 McCluskey⁵³, A. Nazli Basak⁵⁴, Ceren Tunca⁵⁴, Hamid Hamzeiy⁵⁴, Yesim Parman⁵⁵, Thomas
20 Meitinger⁵⁶, Peter Lichtner⁵⁶, Milena Radivojkov-Blagojevic⁵⁶, Christian R. Andres⁵⁷, Cindy
21 Maurel⁵⁷, Gilbert Bensimon^{58,59,60}, Bernhard Landwehrmeyer⁶¹, Alexis Brice^{62,63,64,65,66},
22 Christine A.M. Payan⁶⁷, Safaa Saker-Delye⁶⁸, Alexandra Dürr⁶⁹, Nicholas W. Wood⁷⁰, Lukas
23 Tittmann⁷¹, Wolfgang Lieb⁷¹, Andre Franke⁷², Marcella Rietschel⁷³, Sven Cichon^{74,75,76,77,78},
24 Markus M. Nöthen^{74,75}, Philippe Amouyel⁷⁹, Christophe Tzourio⁸⁰, Jean-François Dartigues⁸⁰,
25 Andre G. Uitterlinden^{81,82}, Fernando Rivadeneira^{81,82}, Karol Estrada⁸¹, Albert Hofman⁸²,
26 Charles Curtis^{6,7}, Hylke M. Blauw¹, Anneke J. van der Kooi⁸³, Marianne de Visser⁸³, An
27 Goris^{84,85}, Markus Weber⁸⁶, Christopher E. Shaw², Bradley N. Smith², Orietta Pansarasa⁸⁷,
28 Cristina Ceredá⁸⁷, Roberto Del Bo⁸⁸, Giacomo P. Comi⁸⁸, Sandra D'Alfonso⁸⁹, Cinzia
29 Bertolin⁹⁰, Gianni Sorarù⁹⁰, Letizia Mazzini⁹¹, Viviana Pensato⁹², Cinzia Gellera⁹², Cinzia
30 Tiloca⁹, Antonia Ratti^{9,14}, Andrea Calvo^{93,94}, Cristina Moglia^{93,94}, Maura Brunetti^{93,94}, Simona
31 Arcuti⁹⁵, Rosa Capozzo⁹⁵, Chiara Zecca⁹⁵, Christian Lunetta⁹⁶, Silvana Penco⁹⁷, Nilo Riva⁹⁸,
32 Alessandro Padovani⁹⁹, Massimiliano Filosto⁹⁹, Bernard Muller¹⁰⁰, Robbert Jan Stuit¹⁰⁰,
33 PARALS registry¹⁰¹, SLALOM group¹⁰¹, SLAP registry¹⁰¹, FALS Sequencing Consortium¹⁰¹,

34 SLAGEN Consortium¹⁰¹, NNIPPS Study Group¹⁰¹, Ian Blair¹⁰², Katharine Zhang¹⁰², Emily P.
35 McCann¹⁰², Jennifer A. Fifita¹⁰², Garth A. Nicholson^{102,103}, Dominic B. Rowe¹⁰², Roger
36 Pamphlett¹⁰⁴, Matthew C. Kiernan¹⁰⁵, Julian Grosskreutz¹⁰⁶, Otto W. Witte¹⁰⁶, Thomas
37 Ringer¹⁰⁶, Tino Prell¹⁰⁶, Beatrice Stubendorff¹⁰⁶, Ingo Kurth¹⁰⁷, Christian A. Hübner¹⁰⁷, P.
38 Nigel Leigh¹⁰⁸, Federico Casale⁹³, Adriano Chio^{93,94}, Ettore Beghi¹⁰⁹, Elisabetta Pupillo¹⁰⁹,
39 Rosanna Tortelli⁹⁵, Giancarlo Logroscino^{110,111}, John Powell², Albert C. Ludolph⁶¹, Jochen H.
40 Weishaupt⁶¹, Wim Robberecht^{85,112,113}, Philip Van Damme^{85,112,113}, Lude Franke⁵, Tune H.
41 Pers^{114,115,116,117,118}, Robert H. Brown¹¹, Jonathan D. Glass^{30,31}, John E. Landers¹¹, Orla
42 Hardiman^{12,13}, Peter M. Andersen^{61,119}, Philippe Corcia^{120,121}, Patrick Vourc'h⁵⁷, Vincenzo
43 Silani^{9,14}, Naomi R. Wray⁸, Peter M. Visscher^{8,122}, Paul I.W. de Bakker^{4,123}, Michael A. van
44 Es¹, R. Jeroen Pasterkamp¹⁰, Cathryn M. Lewis^{6,124}, Gerome Breen^{6,7}, Ammar Al-
45 Chalabi^{2,126,127}, Leonard H. van den Berg^{1,126}, Jan H. Veldink^{1,126,127}

46

Affiliations:

48 1. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center
49 Utrecht, Utrecht, The Netherlands.

50 2. Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and
51 Neuroscience, King's College London, London, UK.

52 3. Population Genetics Laboratory, Smurfit Institute of Genetics, Trinity College Dublin,
53 Dublin, Republic of Ireland.

54 4. Department of Medical Genetics, Center for Molecular Medicine, University Medical
55 Center Utrecht, Utrecht, The Netherlands.

56 5. Department of Genetics, University of Groningen, University Medical Centre Groningen,
57 Groningen, The Netherlands.

58 6. MRC Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry,
59 Psychology & Neuroscience, King's College London, London, UK.

60 7. NIHR Biomedical Research Centre for Mental Health, Maudsley Hospital and Institute of
61 Psychiatry, Psychology & Neuroscience, King's College London, London, UK.

62 8. Queensland Brain Institute, The University of Queensland, Brisbane, Queensland,
63 Australia.

64 9. Department of Neurology and Laboratory of Neuroscience, IRCCS Istituto Auxologico
65 Italiano, Milano, Italy.

66 10. Department of Translational Neuroscience, Brain Center Rudolf Magnus, University
67 Medical Center Utrecht, Utrecht, The Netherlands.

- 68 11. Department of Neurology, University of Massachusetts Medical School, Worcester, MA,
69 USA.
- 70 12. Academic Unit of Neurology, Trinity College Dublin, Trinity Biomedical Sciences
71 Institute, Dublin, Republic of Ireland.
- 72 13. Department of Neurology, Beaumont Hospital, Dublin, Republic of Ireland.
- 73 14. Department of Pathophysiology and Transplantation, 'Dino Ferrari' Center, Università
74 degli Studi di Milano, Milano, Italy.
- 75 15. Department of Biotechnology, Jožef Stefan Institute, Ljubljana, Slovenia.
- 76 16. Biomedical Research Institute BRIS, Ljubljana, Slovenia.
- 77 17. Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine,
78 University of Ljubljana, SI-1000 Ljubljana, Slovenia.
- 79 18. Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, SI-1000
80 Ljubljana, Slovenia.
- 81 19. Ljubljana ALS Centre, Institute of Clinical Neurophysiology, University Medical Centre
82 Ljubljana, SI-1000 Ljubljana, Slovenia.
- 83 20. Institut du Cerveau et de la Moelle épinière, Inserm U1127, CNRS UMR 7225, Sorbonne
84 Universités, UPMC Univ Paris 06 UMRS1127, Paris, France.
- 85 21. Centre de Référence Maladies Rares SLA Ile de France, Département de Neurologie,
86 Hôpital de la Pitié-Salpêtrière, Paris, France.
- 87 22. GRC-UPMC SLA et maladies du Motoneurone, France.
- 88 23. Ramsay Generale de Santé, Hopital Peupliers, Paris, France.
- 89 24. Réseau SLA Ile de France.
- 90 25. Institute of Physiology, Institute of Molecular Medicine, Faculty of Medicine, University
91 of Lisbon, Lisbon, Portugal.
- 92 26. Department of Neurosciences, Hospital de Santa Maria-CHLN, Lisbon, Portugal.
- 93 27. Department of Neurology, Hospital Carlos III, Madrid, Spain.
- 94 28. Neurology Department, Hospital de la Santa Creu i Sant Pau de Barcelona, Autonomous
95 University of Barcelona, Barcelona, Spain.
- 96 29. Centro de Investigación en red en Enfermedades Raras (CIBERER), Spain.
- 97 30. Department Neurology, Emory University School of Medicine, Atlanta, GA, USA.
- 98 31. Emory ALS Center, Emory University School of Medicine, Atlanta, GA, USA.
- 99 32. Euan MacDonald Centre for Motor Neurone Disease Research, Edinburgh, UK.
- 100 33. Centre for Neuroregeneration and Medical Research Council Centre for Regenerative
101 Medicine, University of Edinburgh, Edinburgh, UK.

- 102 34. School of Clinical and Experimental Medicine, College of Medical and Dental Sciences,
103 University of Birmingham, Birmingham, UK.
- 104 35. Queen Elizabeth Hospital, University Hospitals Birmingham NHS Foundation Trust,
105 Birmingham, UK.
- 106 36. Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield,
107 Sheffield, UK.
- 108 37. Department of Molecular Neuroscience, Institute of Neurology, University College
109 London, UK.
- 110 38. Department of Clinical Neuroscience, Institute of Neurology, University College London,
111 UK.
- 112 39. Reta Lila Weston Institute, Institute of Neurology, University College London, UK.
- 113 40. Department of Neurodegenerative Diseases, Institute of Neurology, University College
114 London, UK.
- 115 41. Centre for Neuroscience and Trauma, Blizard Institute, Queen Mary University of
116 London, London, UK.
- 117 42. North-East London and Essex Regional Motor Neuron Disease Care Centre, London, UK.
- 118 43. Department of Neurology, Hannover Medical School, Hannover, Germany.
- 119 44. Department of Neurology, Otto-von-Guericke University Magdeburg, Magdeburg,
120 Germany.
- 121 45. Institute of Clinical Neurobiology, University Hospital Wuerzburg, Germany.
- 122 46. Charité University Hospital, Humboldt-University, Berlin, Germany.
- 123 47. University Medical Center Utrecht, Department of Psychiatry, Rudolf Magnus Institute of
124 Neuroscience, The Netherlands.
- 125 48. Department of Human Genetics, David Geffen School of Medicine, University of
126 California, Los Angeles, CA, USA.
- 127 49. Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human
128 Behavior, University of California, Los Angeles, CA, USA.
- 129 50. Department of Neurology, David Geffen School of Medicine, University of California,
130 Los Angeles, CA, USA.
- 131 51. Department of Neurology, University of California, San Francisco, CA, USA.
- 132 52. Center for Neurodegenerative Disease Research, Perelman School of Medicine at the
133 University of Pennsylvania, Philadelphia, PA, USA.
- 134 53. Department of Neurology, Perelman School of Medicine at the University of
135 Pennsylvania, PA USA.

- 136 54. Neurodegeneration Research Laboratory, Bogazici University, Istanbul, Turkey.
- 137 55. Neurology Department, Istanbul Medical School, Istanbul University, Istanbul, Turkey.
- 138 56. Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany.
- 139 57. INSERM U930, Université François Rabelais, Tours, France.
- 140 58. APHP, Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpêtrière, France.
- 141 59. Université Pierre & Marie Curie, Pharmacologie, Paris VI, Paris, France.
- 142 60. BESPIM, CHU-Nîmes, Nîmes, France.
- 143 61. Department of Neurology, Ulm University, Ulm, Germany.
- 144 62. INSERM U 1127, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
- 145 63. CNRS UMR 7225, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
- 146 64. Sorbonne Universités, Université Pierre et Marie Curie Paris 06 UMRS 1127, Hôpital de
- 147 la Pitié-Salpêtrière, 75013 Paris, France.
- 148 65. Institut du Cerveau et de la Moelle épinière, Hôpital de la Pitié-Salpêtrière, 75013 Paris,
- 149 France.
- 150 66. APHP, Département de Génétique, Hôpital de la Pitié-Salpêtrière, 75013 Paris, France.
- 151 67. Département de Pharmacologie Clinique, Hôpital de la Pitié-Salpêtrière, APHP, UPMC
- 152 Pharmacologie, Paris 6, Paris, BESPIM, CHU-Nîmes, Nîmes, France.
- 153 68. Genethon, CNRS UMR 8587 Evry, France.
- 154 69. Department of Medical Genetics, l'Institut du Cerveau et de la Moelle Épinière, Hoptial
- 155 Salpêtrière, Paris, France.
- 156 70. Department of Neurogenetics, Institute of Neurology, University College London, UK.
- 157 71. PopGen Biobank and Institute of Epidemiology, Christian Albrechts-University Kiel,
- 158 Kiel, Germany.
- 159 72. Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany.
- 160 73. Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health,
- 161 Mannheim, Germany.
- 162 74. Institute of Human Genetics, University of Bonn, Bonn, Germany.
- 163 75. Department of Genomics, Life and Brain Center, Bonn, Germany.
- 164 76. Division of Medical Genetics, University Hospital Basel, University of Basel, Basel,
- 165 Switzerland.
- 166 77. Department of Biomedicine, University of Basel, Basel, Switzerland.
- 167 78. Institute of Neuroscience and Medicine INM-1, Research Center Juelich, Juelich,
- 168 Germany.
- 169 79. University of Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1167 - RID-AGE - Risk

- 170 Factor and molecular determinants of aging diseases, Labex Distalz, F-59000 Lille, France.
171 80. Bordeaux University, ISPED, Centre INSERM U1219-Epidemiologie-Biostatistique &
172 CIC-1401, CHU de Bordeaux, Pole de Sante Publique, Bordeaux, France.
173 81. Department of Internal Medicine, Genetics Laboratory, Erasmus Medical Center
174 Rotterdam, Rotterdam, The Netherlands.
175 82. Department of Epidemiology, Erasmus Medical Center Rotterdam, Rotterdam, The
176 Netherlands.
177 83. Department of Neurology, Academic Medical Center, Amsterdam, The Netherlands.
178 84. Laboratory for Neuroimmunology, Section of Experimental Neurology, Department of
179 Neurosciences, KU Leuven - University of Leuven, Leuven, Belgium.
180 85. Leuven Research Institute for Neuroscience and Disease (LIND), Leuven, Belgium.
181 86. Neuromuscular Diseases Unit/ALS Clinic, Kantonsspital St. Gallen, 9007, St. Gallen,
182 Switzerland.
183 87. Laboratory of Experimental Neurobiology, IRCCS 'C. Mondino' National Institute of
184 Neurology Foundation, Pavia, Italy.
185 88. Neurologic Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Milan,
186 Italy.
187 89. Department of Health Sciences, Interdisciplinary Research Center of Autoimmune
188 Diseases, UPO, Università del Piemonte Orientale, Novara, Italy.
189 90. Department of Neurosciences, University of Padova, Padova, Italy.
190 91. Department of Neurology, University of Eastern Piedmont, Novara, Italy.
191 92. Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS
192 Istituto Neurologico 'Carlo Besta', Milano, Italy.
193 93. "Rita Levi Montalcini" Department of Neuroscience, ALS Centre, University of Torino,
194 Turin, Italy.
195 94. Azienda Ospedaliera Città della Salute e della Scienza, Torino, Italy.
196 95. Department of Clinical research in Neurology, University of Bari "A. Moro", at Pia
197 Fondazione "Card. G. Panico", Tricase (LE), Italy.
198 96. NEMO Clinical Center, Serena Onlus Foundation, Niguarda Ca' Granda Hospital, Milan,
199 Italy.
200 97. Medical Genetics Unit, Department of Laboratory Medicine, Niguarda Ca' Granda
201 Hospital, Milan, Italy.
202 98. Department of Neurology, Institute of Experimental Neurology (INSPE), Division of
203 Neuroscience, San Raffaele Scientific Institute, Milan, Italy.

- 204 99. University Hospital "Spedali Civili", Brescia, Italy.
- 205 100. Project MinE Foundation, Rotterdam, The Netherlands.
- 206 101. A list of members and affiliations appears in the Supplementary Note.
- 207 102. Department of Biomedical Sciences, Faculty of Medicine and Health Sciences,
- 208 Macquarie University, Sydney, New South Wales, Australia.
- 209 103. University of Sydney, ANZAC Research Institute, Concord Hospital, Sydney, New
- 210 South Wales, Australia.
- 211 104. The Stacey MND Laboratory, Department of Pathology, The University of Sydney,
- 212 Sydney, New South Wales, Australia.
- 213 105. Brain and Mind Centre, The University of Sydney, New South Wales 2050, Australia.
- 214 106. Hans-Berger Department of Neurology, Jena University Hospital, Jena, Germany.
- 215 107. Institute of Human Genetics, Jena University Hospital, Jena, Germany.
- 216 108. Department of Neurology, Brighton and Sussex Medical School Trafford Centre for
- 217 Biomedical Research, University of Sussex, Falmer, East Sussex, UK.
- 218 109. Laboratory of Neurological Diseases, Department of Neuroscience, IRCCS Istituto di
- 219 Ricerche Farmacologiche Mario Negri, Milano, Italy.
- 220 110. Department of Basic Medical Sciences, Neuroscience and Sense Organs, University of
- 221 Bari 'Aldo Moro', Bari, Italy.
- 222 111. Unit of Neurodegenerative Diseases, Department of Clinical Research in Neurology,
- 223 University of Bari 'Aldo Moro', at Pia Fondazione Cardinale G. Panico, Tricase, Lecce, Italy.
- 224 112. VIB, Vesalius Research Center, Laboratory of Neurobiology, B-3000 Leuven, Belgium.
- 225 113. University Hospitals Leuven, Department of Neurology, B-3000 Leuven, Belgium.
- 226 114. Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA.
- 227 115. Division of Genetics, Boston Children's Hospital, Boston, MA, USA.
- 228 116. Center for Basic Translational Obesity Research, Boston Children's Hospital, Boston,
- 229 MA, USA.
- 230 117. Department of Genetics, Harvard Medical School, Boston, MA, USA.
- 231 118. Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.
- 232 119. Department of Pharmacology and Clinical Neuroscience, Umeå University, Umeå,
- 233 Sweden.
- 234 120. Centre SLA, CHRU de Tours, Tours, France.
- 235 121. Federation des Centres SLA Tours and Limoges, LITORALS, Tours, France.
- 236 122. Diamantina Institute, The University of Queensland, Translational Research Institute,
- 237 Brisbane, Queensland, Australia.

238 123. Department of Epidemiology, Julius Center for Health Sciences and Primary Care,
239 University Medical Center Utrecht, Utrecht, The Netherlands.

240 124. Department of Medical and Molecular Genetics, King's College London, London, UK.

241 125. These authors contributed equally to this work.

242 126. These authors jointly directed this work.

243 127. To whom correspondence should be addressed.

244

245 **Corresponding authors:**

246 Ammar Al-Chalabi

247 Department of Basic and Clinical Neuroscience

248 King's College London

249 Maurice Wohl Clinical Neuroscience Institute

250 Coldharbour Lane

251 E-mail: ammar.al-chalabi@kcl.ac.uk

252

253 Jan H. Veldink

254 Department of Neurology

255 Brain Center Rudolf Magnus

256 University Medical Center Utrecht

257 Heidelberglaan 100

258 3584CX Utrecht

259 The Netherlands

260 Email: j.h.veldink@umcutrecht.nl

261

262 **To elucidate the genetic architecture of amyotrophic lateral sclerosis (ALS) and find**
263 **associated loci, we assembled a custom imputation reference panel from whole genome-**
264 **sequenced ALS patients and matched controls ($N = 1,861$). Through imputation and**
265 **mixed-model association analysis in 12,577 cases and 23,475 controls, combined with**
266 **2,579 cases and 2,767 controls in an independent replication cohort, we fine mapped a**
267 **novel locus on chromosome 21 and identified *C21orf2* as an ALS risk gene. In addition,**
268 **we identified *MOBP* and *SCFD1* as novel associated risk loci. We established evidence**
269 **for ALS being a complex genetic trait with a polygenic architecture. Furthermore, we**
270 **estimated the SNP-based heritability at 8.5%, with a distinct and important role for low**
271 **frequency (1–10%) variants. This study motivates the interrogation of larger sample**
272 **sizes with full genome coverage to identify rare causal variants that underpin ALS risk.**
273

274 ALS is a fatal neurodegenerative disease that affects 1 in 400 people, death occurring within
275 three to five years¹. Twin-based studies estimate heritability to be around 65% and 5–10% of
276 ALS patients have a positive family history^{1,2}. Both are indicative of an important genetic
277 component in ALS etiology. Following the initial discovery of the *C9orf72* locus in GWASs^{3–}
278 ⁵, the identification of the pathogenic hexanucleotide repeat expansion in this locus
279 revolutionized the field of ALS genetics and biology^{6,7}. The majority of ALS heritability,
280 however, remains unexplained and only two additional risk loci have been identified robustly
281 since^{3,8}.

282
283 To discover new genetic risk loci and elucidate the genetic architecture of ALS, we genotyped
284 7,763 new cases and 4,669 controls and additionally collected existing genotype data of
285 published GWAS in ALS. In total, we analyzed 14,791 cases and 26,898 controls from 41
286 cohorts (**Supplementary Table 1, Supplementary Methods**). We combined these cohorts
287 based on genotyping platform and nationality to form 27 case-control strata. In total 12,577
288 cases and 23,475 controls passed quality control (Online methods, **Supplementary Tables 2–**
289 **5**).

290
291 For imputation purposes we obtained high-coverage (~43.7X) whole genome sequencing data
292 from 1,246 ALS patients and 615 controls from The Netherlands (Online methods, and
293 **Supplementary Fig. 1**). After quality control, we constructed a reference panel including
294 18,741,510 single nucleotide variants. Imputing this custom reference panel into Dutch ALS
295 cases increased imputation accuracy of low-frequency genetic variation (minor allele

296 frequency, MAF 0.5–10%) considerably compared to commonly used reference panels: the
297 1000 Genomes Project phase 1 (1000GP)⁹ and Genome of The Netherlands (GoNL)¹⁰ (**Fig.**
298 **1a**). The improvement was also observed when this reference panel was used to impute into
299 ALS cases from the UK (**Fig. 1b**). To benefit from the global diversity of haplotypes, the
300 custom and 1000GP panels were combined, which further improved imputation. Given these
301 results, we used the merged reference panel for imputation of all strata in our study.

302
303 In total we imputed 8,697,640 variants passing quality control in the 27 strata and separately
304 tested these for association with ALS risk by logistic regression. Results were then included
305 in an inverse-variance weighted fixed effects meta-analysis, which revealed 4 loci at genome-
306 wide significance ($p < 5 \times 10^{-8}$) (**Fig. 2a**). The previously reported *C9orf72* (rs3849943)^{3–5,8},
307 *UNC13A* (rs12608932)^{3,5} and *SARMI* (rs35714695)⁸ loci all reached genome-wide
308 significance, as did a novel association for a non-synonymous variant in *C21orf2*
309 (rs75087725, $p = 8.7 \times 10^{-11}$, **Supplementary Tables 8 and 10–13**). Interestingly, this variant
310 was present on only 10 haplotypes in the 1000GP reference panel (MAF = 1.3%), while our
311 custom reference panel included 62 haplotypes carrying the minor allele (MAF = 1.7%). As a
312 result, more strata passed quality control for this variant by passing the allele frequency
313 threshold of 1% (**Supplementary Table 9**). This demonstrates the benefit of the merged
314 reference panel with ALS-specific content, which improved imputation and resulted in a
315 genome-wide significant association.

316
317 Linear mixed models (LMM) can improve power while controlling for sample structure¹¹,
318 particularly in our study that included a large number of imperfectly balanced strata. Even
319 though LMM for ascertained case-control data has a potential small loss of power¹¹, we
320 judged the advantage of combining all strata while controlling the false positive rate, to be
321 more important and therefore jointly analyzed all strata in a LMM to identify additional risk
322 loci. There was no overall inflation of the linear mixed model's test statistic compared to the
323 meta-analysis (**Supplementary Fig. 2**). We observed modest inflation in the QQ-plot ($\lambda_{GC} =$
324 1.12, $\lambda_{1000} = 1.01$, **Supplementary Fig. 3**). LD score regression yielded an intercept of 1.10
325 (standard error 7.8×10^{-3}). While the LD score regression intercept can indicate residual
326 population stratification, which is fully corrected for in a LMM, the intercept can also reflect
327 a distinct genetic architecture where most causal variants are rare, or a non-infinitesimal
328 architecture¹². The linear mixed model identified all four genome-wide significant

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

339
340 In the replication phase, we genotyped the newly discovered associated SNPs in nine
341 independent replication cohorts, totaling 2,579 cases and 2,767 controls. In these cohorts we
342 replicated the signals for the *C21orf2*, *MOBP* and *SCFDI* loci, with lower p-values in the
343 combined analysis than the discovery phase (combined p-value = 3.08×10^{-10} , $p = 4.19 \times 10^{-$
344 10 and $p = 3.45 \times 10^{-8}$ for rs75087725, rs616147 and rs10139154 respectively, **Table 1,**
345 **Supplementary Fig. 4**)¹⁷. The combined signal for rs7813314 was less significant due to an
346 opposite effect between the discovery and replication phase, indicating non-replication.
347 Although replication yielded similar effect estimates for rs10139154 compared to the
348 discovery phase, this was not statistically significant ($p = 0.09$) in the replication phase alone.
349 This reflects the limited sample size of our replication phase, which is inherent to the low
350 prevalence of ALS and warrants even larger sample sizes to replicate this signal robustly.

351
352 There was no evidence for residual association within each locus after conditioning on the top
353 SNP, indicating that all risk loci are independent signals. Apart from the *C9orf72*, *UNC13A*
354 and *SARMI* loci, we found no evidence for associations previously described in smaller
355 GWAS (**Supplementary Table 17**).

356
357 The associated low-frequency non-synonymous SNP in *C21orf2* suggested that this gene
358 could directly be involved in ALS risk. Indeed, we found no evidence that linkage
359 disequilibrium of sequenced variants beyond *C21orf2* explained the association within this
360 locus (**Supplementary Fig. 5**). In addition, we investigated the burden of rare coding
361 mutations in a set of whole genome sequenced cases (N = 2,562) and controls (N = 1,138).
362 After quality control these variants were tested using a pooled association test for rare variants

363 corrected for population structure (T5 and T1 for 5% and 1% allele frequency,
364 **Supplementary methods**). This revealed an excess of non-synonymous and loss-of-function
365 mutations in *C21orf2* among ALS cases that persists after conditioning on rs75087725 ($p_{T5} =$
366 9.2×10^{-5} , $p_{T1} = 0.01$, **Supplementary Fig. 6**), which further supports that *C21orf2*
367 contributes to ALS risk.

368

369 In an effort to fine-map the other loci to susceptibility genes, we searched for SNPs in these
370 loci with *cis*-eQTL effects observed in brain and other tissues (**Supplementary methods**,
371 **Supplementary Table 18**)¹⁸. There was overlap with previously identified brain *cis*-eQTLs
372 for five regions (**Supplementary Fig. 7, Supplementary Table 19, Supplementary Data**
373 **Set 1**). Interestingly, within the *C9orf72* locus we found that proxies of rs3849943 (LD $r^2 =$
374 $0.21 - 0.56$) had a brain *cis*-eQTL effect on *C9orf72* only (minimal $p = 5.27 \times 10^{-7}$), which
375 harbors the hexanucleotide repeat expansion that drives this GWAS signal. Additionally, we
376 found that rs12608932 and its proxies within the *UNC13A* locus had exon-level *cis*-eQTL
377 effect on *KCNN1* in frontal cortex ($p = 1.15 \times 10^{-3}$)¹⁹. Another overlap was observed in the
378 *SARM1* locus where rs35714695 and its proxies had the strongest exon-level *cis*-eQTL effect
379 on *POLDIP2* in multiple brain tissues ($p = 2.32 \times 10^{-3}$). Within the *SCFD1* locus rs10139154
380 and proxies had a *cis*-eQTL effect on *SCFD1* in cerebellar tissue ($p = 7.71 \times 10^{-4}$). For the
381 *MOBP* locus, rs1768208 and proxies had a *cis*-eQTL effect on *RPSA* ($p = 7.71 \times 10^{-4}$).

382

383 To describe the genetic architecture of ALS, we calculated polygenic scores that can be used
384 to predict phenotypes for traits with a polygenic architecture²⁰. We calculated the SNP effects
385 using a linear mixed model in 18 of the 27 strata and subsequently assessed their predictive
386 ability in the other 9 independent strata. The analysis revealed that a significant, albeit
387 modest, proportion of the phenotypic variance could be explained by all SNPs (Nagelkerke r^2
388 $= 0.44\%$, $r^2 = 0.15\%$ on the liability scale, $p = 2.7 \times 10^{-10}$, **Supplementary Fig. 8**). This
389 finding adds to the existing evidence that ALS is a complex genetic trait with a polygenic
390 architecture. To further quantify the contribution of common SNPs to ALS risk, we estimated
391 the SNP-based heritability using three approaches, all assuming a population baseline risk of
392 0.25% ²¹. The variance explained by all SNPs using GCTA-REML estimated heritability at
393 8.5% (SE 0.5%). Haseman-Elston regression yielded a very similar 7.9% and LD score
394 regression estimated the SNP-based heritability at 8.2% (SE 0.5%). The heritability estimates
395 per chromosome were strongly correlated with chromosome length ($p = 4.9 \times 10^{-4}$, $r^2 = 0.46$,
396 **Fig. 3a**), which again is indicative of the polygenic architecture of ALS.

397

398 We found that the genome-wide significant loci only explained 0.2% of the heritability and
399 thus the bulk of the heritability (8.3%, SE 0.3%) was captured in SNPs below genome-wide
400 significance. This implies that many genetic risk variants have yet to be discovered.

401 Understanding where these unidentified risk variants remain across the allele frequency
402 spectrum will inform designing future studies to identify these variants. We, therefore,
403 estimated heritability partitioned by minor allele frequency. Furthermore, we contrasted this
404 to common polygenic traits studied in GWASs such as schizophrenia. We observed a clear
405 trend that indicated that most variance is explained by low-frequency SNPs (**Fig. 3b**).

406 Exclusion of the *C9orf72* locus, which harbors the rare pathogenic repeat expansion, and the
407 other genome-wide significant loci did not affect this trend (**Supplementary fig. 9**). This
408 architecture is different from that expected for common polygenic traits and reflects a
409 polygenic rare-variant architecture observed in simulations²².

410

411 To gain better insight into the biological pathways that explain the associated loci found in
412 this study we looked for enriched pathways using DEPICT²³. This revealed SNAP receptor
413 (SNARE) activity as the only enriched category (FDR < 0.05, **Supplementary Fig. 10**).
414 SNARE complexes play a central role in neurotransmitter release and synaptic function²⁴,
415 which are both perturbed in ALS²⁵.

416

417 Although the biological role of *C21orf2*, a conserved leucine-rich repeat protein, remains
418 poorly characterized, it is part of the ciliome and is required for the formation and/or
419 maintenance of primary cilia²⁶. Defects in primary cilia are associated with various
420 neurological disorders and cilia numbers are decreased in G93A *SOD1* mice, a well-
421 characterized ALS model²⁷. *C21orf2* has also been localized to mitochondria in immune
422 cells²⁸ and is part of the interactome of the protein product of *NEK1*, which has previously
423 been associated with ALS¹⁵. Both proteins appear to be involved in DNA repair
424 mechanisms²⁹. Although future studies are needed to dissect the function of *C21orf2* in ALS
425 pathophysiology it is tempting to speculate that defects in *C21orf2* lead to primary cilium
426 and/or mitochondrial dysfunction or inefficient DNA repair and thereby adult onset disease.
427 The other associated loci will require more extensive studies to fine-map causal variants. The
428 *SARM1* gene has been suggested as a susceptibility gene for ALS, mainly because of its role
429 in Wallerian degeneration and interaction with *UNC13A*^{8,30}. Although these are indeed
430 interesting observations, the brain *cis*-eQTL effect on *POLDIP2* suggests that *POLDIP2* and

431 not *SARM1* could in fact be the causal gene within this locus. Similarly, *KCNN1*, which
432 encodes a neuronal potassium channel involved in neuronal excitability, could be the causal
433 gene either through a direct eQTL effect or rare variants in LD with the associated SNP in
434 *UNC13A*.

435

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

441 **ACCESSION CODES**

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

447

448 **DATA ACCESS**

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

451

452 **ACKNOWLEDGMENTS**

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

455

456 **AUTHOR CONTRIBUTIONS**

457 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,
458 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.,
459 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.,
460 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.,
461 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.,
462 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.,
463 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.,
464 V.P., C.G., C.Tiloca, A.R., A.Calvo., C.Moglia, M.B., S.Arcuti., R.C., C.Z., C.L., S.Penco,
465 N.R., A.Padovani., M.F., B.M., R.J.S., PARALS registry, SLALOM group, SLAP registry,
466 FALS sequencing consortium, SLAGEN consortium, NNIPPS study group, I.B., G.A.N.,
467 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.Chìò.,
468 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.,
469 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
470 phenotyping, sample collection and management. W.v.R., A.S., A.M.D., R.L.M., F.P.D.,
471 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.,
472 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,
473 G.B., A.A.-C and J.H.V. prepared DNA and performed SNP-array hybridizations. W.v.R.,
474 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

475 involved in the next-generations sequencing analyses. W.v.R., K.R.v.E., A.M., P.I.W.d.B.,
476 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.,
477 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.,
478 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.,
479 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.
480 performed the replication analyses. W.v.R., A.S., R.L.M., M.R.R., J.Y., N.R.W., P.M.V.,
481 C.L., A.A.-C and J.H.V. performed polygenic risk scoring and heritability analyses. S.d.J.,
482 U.V., L.F., T.P., W.v.R., O.H., G.B., R.J.P. and J.H.V. performed biological pathway
483 analyses. U.V., L.F., W.v.R. and J.H.V. performed eQTL analyses. W.v.R., A.S., A.A.-C.,
484 L.H.v.d.B. and J.H.V., prepared the manuscript with contributions from all authors. A.A.-C.,
485 L.H.v.d.B. and J.H.V. directed the study.

486

487 **REFERENCES FOR MAIN TEXT**

- 488 1. Hardiman, O., van den Berg, L. H. & Kiernan, M. C. Clinical diagnosis and
489 management of amyotrophic lateral sclerosis. *Nat Rev Neurol* **7**, 639–649 (2011).
- 490 2. Al-Chalabi, A. *et al.* An estimate of amyotrophic lateral sclerosis heritability using
491 twin data. *J. Neurol. Neurosurg. Psychiatry* **81**, 1324–1326 (2010).
- 492 3. van Es, M. A. *et al.* Genome-wide association study identifies 19p13.3 (UNC13A) and
493 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat. Genet.* **41**, 1083–
494 1087 (2009).
- 495 4. Laaksovirta, H. *et al.* Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a
496 genome-wide association study. *Lancet Neurol.* **9**, 978–985 (2010).
- 497 5. Shatunov, A. *et al.* Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the
498 UK and seven other countries: a genome-wide association study. *Lancet Neurol.* **9**, 986–994
499 (2010).
- 500 6. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in
501 noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**,
502 245–256 (2011).
- 503 7. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of
504 chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
- 505 8. Fogh, I. *et al.* A genome-wide association meta-analysis identifies a novel locus at
506 17q11.2 associated with sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **23**, 2220–
507 2231 (2014).

- 508 9. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from
509 1,092 human genomes. *Nature* **491**, 56–65 (2012).
- 510 10. Genome of the Netherlands Consortium. Whole-genome sequence variation,
511 population structure and demographic history of the Dutch population. *Nat. Genet.* **46**, 818–
512 825 (2014).
- 513 11. Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M. & Price, A. L. Advantages
514 and pitfalls in the application of mixed-model association methods. *Nat. Genet.* **46**, 100–106
515 (2014).
- 516 12. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from
517 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 518 13. Höglinger, G. U. *et al.* Identification of common variants influencing risk of the
519 tauopathy progressive supranuclear palsy. *Nat. Genet.* (2011).
- 520 14. Irwin, D. J. *et al.* Myelin oligodendrocyte basic protein and prognosis in behavioral-
521 variant frontotemporal dementia. *Neurology* **83**, 502–509 (2014).
- 522 15. Cirulli, E. T. *et al.* Exome sequencing in amyotrophic lateral sclerosis identifies risk
523 genes and pathways. *Science* **347**, 1436–1441 (2015).
- 524 16. Freischmidt, A. *et al.* Haploinsufficiency of TBK1 causes familial ALS and fronto-
525 temporal dementia. *Nat. Neurosci.* **18**, 631–636 (2015).
- 526 17. Skol, A. D., Scott, L. J., Abecasis, G. R. & Boehnke, M. Joint analysis is more
527 efficient than replication-based analysis for two-stage genome-wide association studies. *Nat.*
528 *Genet.* **38**, 209–213 (2006).
- 529 18. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to
530 enhance discovery from GWAS. *PLoS Genet.* **6**, (2010).
- 531 19. Ramasamy, A. *et al.* Genetic variability in the regulation of gene expression in ten
532 regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014)
- 533 20. Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**,
534 507–515 (2013).
- 535 21. Johnston, C. A. *et al.* Amyotrophic lateral sclerosis in an urban setting: a population
536 based study of inner city London. *J. Neurol.* **253**, 1642–1643 (2006).
- 537 22. Lee, S. H. *et al.* Estimating the proportion of variation in susceptibility to
538 schizophrenia captured by common SNPs. *Nat. Genet.* **44**, 247–250 (2012).
- 539 23. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using
540 predicted gene functions. *Nat. Commun.* **6**, 5890–20 (2015).

- 541 24. Ramakrishnan, N. A., Drescher, M. J. & Drescher, D. G. The SNARE complex in
542 neuronal and sensory cells. *Mol. Cell. Neurosci.* **50**, 58–69 (2012).
- 543 25. Ferraiuolo, L., Kirby, J., Grierson, A. J., Sendtner, M. & Shaw, P. J. Molecular
544 pathways of motor neuron injury in amyotrophic lateral sclerosis. *Nat. Rev. Neurol.* **7**, 616–
545 630 (2011).
- 546 26. Lai, C. K. *et al.* Functional characterization of putative cilia genes by high-content
547 analysis. *Mol. Biol. Cell.* **22**, 1104–1119 (2011).
- 548 27. Ma, X., Peterson, R. & Turnbull, J. Adenylyl cyclase type 3, a marker of primary cilia,
549 is reduced in primary cell culture and in lumbar spinal cord in situ in G93A SOD1 mice.
550 *BMC. Neurosci.* **12**, 71 (2011).
- 551 28. Krohn, K., Ovod, V., Vilja, P., Heino, M. & Scott, H. Immunochemical
552 characterization of a novel mitochondrially located protein encoded by a nuclear gene within
553 the DFNB8/10 critical region on 21q22. 3. *Biochem. Biophys. Res. Commun.* (1997).
- 554 29. Fang, X. *et al.* The NEK1 interactor, C21ORF2, is required for efficient DNA damage
555 repair. *Acta Biochim. Biophys. Sin.* **47**, 834–841 (2015)
- 556 30. Vérièpe, J., Fossouo, L. & Parker, J. A. Neurodegeneration in *C. elegans* models of
557 ALS requires TIR-1/Sarm1 immune pathway activation in neurons. *Nat. Commun.* **6**, 7319
558 (2015).

559

560 **FIGURE LEGENDS**

561 **Figure 1. Imputation accuracy comparison.** The aggregate r^2 value between imputed and
562 sequenced genotypes on chromosome 20 using different reference panels for imputation.
563 Allele frequencies are calculated from the Dutch samples included in the Genome of the
564 Netherlands cohort. The highest imputation accuracy was achieved when imputing from the
565 merged custom and 1000GP panels. This difference is most pronounced for low frequency
566 (0.5–10%) alleles in both ALS cases from The Netherlands (a) and United Kingdom (b).

567

568 **Figure 2. Meta-analysis and linear mixed model associations.** (a) Manhattan plot for meta-
569 analysis results. This yielded four genome-wide significant associations highlighted with
570 names indicating the closest gene. The associated SNP in *C21orf2* is a non-synonymous
571 variant not found in previous GWAS. (b) Manhattan plot for linear mixed model results. This
572 association analysis yielded three additional loci reaching genome-wide significance (*MOBP*,
573 *LOC101927815* and *SCFD1*). SNPs in the previously identified ALS risk gene *TBKI*
574 approached genome-wide significance ($p = 6.6 \times 10^{-8}$). Since the *C21orf2* SNP was removed

575 from a Swedish stratum because of a $MAF < 1\%$, this SNP was tested separately, but is
 576 presented here together with all other SNPs with a $MAF > 1\%$ in every stratum. Here,
 577 *LOC101927815* is colored grey because the association for this locus could not be replicated.

578

579 **Figure 3. Partitioned heritability.** (a) The heritability estimates per chromosome were
 580 strongly correlated with chromosome length ($p = 4.9 \times 10^{-4}$). (b) For ALS there was a clear
 581 trend where more heritability was explained within the lower allele frequency bins. This
 582 effect was still observed when, for a fair comparison between ALS and a previous study
 583 partitioning heritability for schizophrenia (SCZ) using identical methods²², SNPs present in
 584 HapMap3 (HM3) were included. The pattern for ALS resembles that observed in a rare
 585 variant model simulation performed in this study. Error bars reflect standard errors.

586

587 **TABLES**588 **Table 1. Discovery and replication of novel genome-wide significant loci.**

SNP	Discovery					Replication				Combined	
	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**

589

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

600

601 **AUTHOR INFORMATION**

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

605

606 **ONLINE METHODS**

607 Software packages used, their version, web source, and references are described in the
608 **Supplementary Table 20**.

609

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

616 Subsequent quality control included:

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

622 More details are described in the **Supplementary methods**. The number of SNPs and
623 individuals failing each QC step per cohort and stratum are displayed in **Supplementary**
624 **Tables 2–5**.

625

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

632

633 **Merging reference panels.** All high quality calls in the custom reference panel were phased
634 using SHAPEIT2 software. After checking strand and allele inconsistencies, both the 1000
635 Genomes Project (1000GP) reference panel (release 05-21-2011)³¹ and custom reference

636 panel were imputed up to the union of their variants as described previously³². Those variants
637 with inconsistent allele frequencies between the two panels were removed.

638
639 **Imputation accuracy performance.** To assess the imputation accuracy between different
640 reference panels, 109 unrelated ALS cases of Dutch ancestry sequenced by Complete
641 Genomics and 67 ALS cases from the UK sequenced by Illumina were selected as a test
642 panel. All variants not present on the Illumina Omni1M array were masked and the SNVs on
643 chromosome 20 were subsequently imputed back using four different reference panels
644 (1000GP, GoNL, custom panel and merged panel). Concordance between the imputed alleles
645 and sequenced alleles was assessed within each allele frequency bin where allele frequencies
646 are calculated from the Dutch samples included in the Genome of the Netherlands cohort.

647
648 **GWAS imputation.** Pre-phasing was performed per stratum using SHAPEIT2 with the
649 1000GP phase 1 (release 05-21-2011) haplotypes³¹ as a reference panel. Subsequently, strata
650 were imputed up to the merged reference panel in 5 megabase chunks using IMPUTE2.
651 Imputed variants with a MAF < 1% or INFO score < 0.3 were excluded from further analysis.
652 Variants with allele frequency differences between strata, defined as deviating > 10SD from
653 the normalized mean allele frequency difference between those strata and an absolute
654 difference > 5%, were excluded, since they are likely to represent sequencing or genotyping
655 artifacts. Imputation concordance scores for cases and controls were compared to assess
656 biases in imputation accuracy (**Supplementary Table 6**).

657
658 **Meta-analysis.** Logistic regression was performed on imputed genotype dosages under an
659 additive model using SNPTTEST software. Based on scree plots, 1 to 4 principal components
660 were included per stratum. These results were then combined in an inverse-variance weighted
661 fixed effect meta-analysis using METAL. No marked heterogeneity across strata was
662 observed as the Cochran's Q test statistics did not deviate from the null-distribution ($\lambda =$
663 0.96). Therefore, no SNPs were removed due to excessive heterogeneity. The genomic
664 inflation factor was calculated and the quantile-quantile plot is provided in **Supplementary**
665 **Fig. 3a**.

666
667 **Linear mixed model.** All strata were combined including SNPs that passed quality control in
668 every stratum. Subsequently the genetic relationship matrices (GRM) were calculated per

669 chromosome including all SNPs using the Genome-Wide Complex Trait Analysis (GCTA)
670 software package. Each SNP was then tested in a linear mixed model including a GRM
671 composed of all chromosomes excluding the target chromosome (leave one chromosome out,
672 LOCO). The genomic inflation factor was calculated and the quantile-quantile plot is
673 provided as **Supplementary Fig. 3b**.

674
675 **Replication.** For the replication phase independent ALS cases and controls from Australia,
676 Belgium, France, Germany, Ireland, Italy, The Netherlands and Turkey that were not used in
677 the discovery phase were included. A pre-designed TaqMan genotyping assay was used to
678 replicate rs75087725 and rs616147. Sanger sequencing was performed to replicate
679 rs10139154 and rs7813314 (**Supplementary methods, Supplementary Table 7**). All
680 genotypes were tested in a logistic regression per country and subsequently meta-analyzed.

681
682 **Rare variant analysis in *C21orf2*.** The burden of non-synonymous rare variants in *C21orf2*
683 was assessed in whole genome sequencing data obtained from ALS cases and controls from
684 The Netherlands, Belgium, Ireland, United Kingdom and the United States. After quality
685 control the burden of non-synonymous and loss-of-function mutations in *C21orf2* were tested
686 for association per country and subsequently meta-analyzed. More details are provided in the
687 **Supplementary methods**.

688
689 **Polygenic risk scores.** To assess the predictive accuracy of polygenic risk scores in an
690 independent dataset SNP weights were assigned based on the linear mixed model (GCTA-
691 LOCO) analysis in 18/27 strata. SNPs in high LD ($r^2 > 0.5$) within a 250 kb window were
692 clumped. Subsequently, polygenic risk scores for cases and controls in the 9 independent
693 strata were calculated based on their genotype dosages using PLINK v1.9. To obtain the
694 Nagelkerke R^2 and corresponding p-values these scores were then regressed on their true
695 phenotype in a logistic regression where (based on scree plots) the first three PCs, sex and
696 stratum were included as covariates.

697
698 **SNP-based heritability estimates. GCTA-REML.** GRMs were calculated using GCTA
699 software including genotype dosages passing quality control in all strata. Based on the
700 diagonal of the GRM individuals representing subpopulations that contain an abundance of
701 rare alleles (diagonal values mean \pm 2SD) were removed (**Supplementary Fig. 14a**). Pairs
702 where relatedness (off-diagonal) exceeded 0.05 were removed as well (**Supplementary Fig.**

703 **14b**). The eigenvectors for the first 10 PCs were included as fixed effects to account for more
704 subtle population structure. The prevalence of ALS was defined as the life-time morbid risk
705 for ALS (i.e. 1/400)¹⁹. To estimate the SNP-based heritability for all non-genome-wide
706 significant SNPs, genotypes for the SNPs reaching genome-wide significance were modeled
707 as fixed effect. The variance explained by the GRM therefore reflects the SNP-based
708 heritability of all non-genome-wide significant SNPs. SNP-based heritability partitioned by
709 chromosome or MAF was calculated by including multiple GRMs, calculated on SNPs from
710 each chromosome or within the respective frequency bin, in one model.

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

715 *LD score regression*. Summary statistics from GCTA-LOCO and LD scores calculated from
716 European individuals in 1000GP were used for LD score regression. Strongly associated
717 SNPs ($p < 5 \times 10^{-8}$) and variants not in HapMap3 were excluded. Considering adequate
718 correction for population structure and distant relatedness in the linear mixed model, the
719 intercept was constrained to 1.0¹².

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

727

728 REFERENCES FOR METHODS

- 729 31. Delaneau, O., Marchini, J. & 1000 Genomes Project Consortium. Integrating sequence
730 and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat.*
731 *Commun.* **5**, 3934 (2014).
- 732 32. Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of
733 genomes. *G3* **1**, 457-70 (2011).