Genomic analysis of intracranial and subcortical brain volumes yields polygenic scores accounting for variation across ancestries

Luis M García-Marín[1,2]*, Adrian I Campos[1,3]*, Santiago Diaz-Torres[2,4], Jill A Rabinowitz[5], Zuriel Ceja[1,2], Brittany L Mitchell[1,2], Katrina L Grasby[1,2], Jackson G Thorp[1], Ingrid Agartz[6–8], Saud Alhusaini[9,10], David Ames[11,12], Philippe Amouyel[13–16], Ole A Andreassen[6,17,18], Konstantinos Arfanakis[19,20], Alejandro Arias Vasquez[21], Nicola J Armstrong[22], Lavinia Athanasiu[6,23], Mark E Bastin[24], Alexa S Beiser[25,26], David A Bennett[19,27], Joshua C Bis[28], Marco PM Boks[29], Dorret I Boomsma[30], Henry Brodaty[31], Rachel M Brouwer[32], Jan K Buitelaar[33], Ralph Burkhardt[34,35], Wiepke Cahn[36,37], Vince D. Calhoun[38], Owen T Carmichael[39], Mallar Chakravarty[40,41], Qiang Chen[42], Christopher R. K. Ching[43], Sven Cichon[44–46], Benedicto Crespo-Facorro[47], Fabrice Crivello[48], Anders M Dale[49], George Davey Smith[50,51], Eco JC de Geus[52], Philip L. De Jager[53], Greig I de Zubicaray[54], Stéphanie Debette[54,56], Charles DeCarli[57], Chantal Depondt[58], Sylvane Desrivières[59], Srdjan Djurovic[60,61], Stefan Ehrlich[62], Susanne Erk[63], Thomas Espeseth[64,65], Guillén Fernández[66], Irina Filippi[67], Simon E Fisher[68,69], Debra A Fleischman[19,70], Evan Fletcher[71], Myriam Fornage[72], Andreas J Forstner[44,73], Clyde Francks[66,68,74], Barbara Franke[33,75], Tian Ge[76], Aaron L Goldman[42], Hans J Grabe[77], Robert C Green[78], Oliver Grimm[79,80], Nynke A Groenewold[81], Oliver Gruber[82], Vilmundur Gudnason[83,84], Asta K Håberg[85,86], Unn K Haukvik[87,88], Andreas Heinz[88], Derrek P Hibar[90], Saima Hilal[91], Jayandra J Himali[25,26,92-94], Beng-Choon Ho[95], David F Hoehn[96], Pieter J Hoekstra[97,98], Edith Hofer[99,100], Wolfgang Hoffmann[101,102], Avram J Holmes[103], Georg Homuth[104], Norbert Hosten[105], M. Kamran Ikram[106], Jonathan C Ipser[107], Clifford R Jack Jr[108], Neda Jahanshad[43], Erik G Jönsson[6,109], Rene S Kahn[36], Ryota Kanai[110], Marieke Klein[66,75], Maria J Knol[111], Lenore J Launer[112], Stephen M Lawrie[113], Stephanie Le Hellard[61], Phil H Lee[114–116], Hervé Lemaître[117], Shuo Li[25,26], David CM Liewald[118], Honghuang Lin[119], W T Longstreth Jr[120,121], Oscar L Lopez[122], Michelle Luciano[123], Pauline Maillard[71], Andre F Marquand[66], Nicholas G Martin[1], Jean-Luc Martinot[124], Karen A Mather[31], Venkata S Mattay[42], Katie L McMahon[125], Patrizia Mecocci[126,127], Ingrid Melle[6], Andreas Meyer-Lindenberg[128], Nazanin Mirza-Schreiber[129,130], Yuri Milaneschi[131–134], Thomas H Mosley[135], Thomas W Mühleisen[44,136,137], Bertram Müller-Myhsok[138], Susana Muñoz Maniega[24], Matthias Nauck[139,140], Kwangsik Nho[141], Wiro J Niessen[142], Markus M Nöthen[73], Paul A Nyquist[143,144], Jaap Oosterlaan[145–147], Massimo Pandolfo[148,149], Tomas Paus[150,151], Zdenka Pausova[152,153], Brenda WJH Penninx[131], G. Bruce Pike[154], Bruce M Psaty[28,121,155], Benno Pütz[156], Simone Reppermund[31,157], Marcella D Rietschel[158], Shannon L Risacher[159,160], Nina Romanczuk-Seiferth[161,162], Rafael Romero-Garcia[163,164], Gennady V Roshchupkin[111,165], Jerome I Rotter[166], Perminder S Sachdev[31,167], Philipp G Sämann[96], Arvin Saremi[43], Muralidharan Sargurupremraj[55,92], Andrew J Saykin[159], Lianne Schmaal[168,169], Helena Schmidt[170], Reinhold Schmidt[171], Peter R Schofield[172,173], Markus Scholz[35,174], Gunter Schumann[175,176], Emanuel Schwarz[128], Li Shen[177], Jean Shin[178], Sanjay M Sisodiya[179,180], Albert V Smith[83,181], Jordan W Smoller[76], Hilkka S Soininen[182], Vidar M Steen[61,183], Dan J Stein[184], Jason L Stein[185], Sophia I Thomopoulos[43], Arthur W. Toga[43], Diana Tordesillas-Gutiérrez[186,187], Julian N Trollor[31,188], Maria C Valdes-Hernandez[24], Dennis van 't Ent[189], Hans van Bokhoven[33,190], Dennis van der Meer[6,191], Nic JA van der Wee[192], Javier Vázquez-Bourgon[193–195], Dick J Veltman[131], Meike W Vernooij[111,165], Arno Villringer[196,197], Louis N Vinke[198], Henry Völzke[102], Henrik Walter[63], Joanna M Wardlaw[24,199], Daniel R Weinberger[42,143,200–202], Michael W Weiner[203–205], Wei Wen[31], Lars

T Westlye[6,64], Eric Westman[206], Tonya White[207], A. Veronica Witte[196,197], Christiane Wolf[96], Jingyun Yang[19,27], Marcel P Zwiers[33], M Arfan Ikram[111], Sudha Seshadri[26,92], Paul M Thompson[43], Claudia L Satizabal[26,208,209], Sarah E Medland[1,2, 54, 210], Miguel E Rentería[1,2]

1. Brain & Mental Health Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, 4006, Australia.
2. School of Biomedical Sciences, Faculty of Medicine, The University of Queensland, Brisbane, QLD, 4072, Australia.
3. Institute for Molecular Biosciences, The University of Queensland, Brisbane, QLD, 4072, Australia.
4. Population Health Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, 4006, Australia.
5. Department of Psychiatry, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ 08854, USA.
6. Centre for Precision Psychiatry, Institute of Clinical Medicine, University of Oslo, Oslo, 0319, Norway.
7. Department of Psychiatric Research, Diakonhjemmet Hospital, Oslo, 0407, Norway.
8. Department of Clinical Neuroscience, Karolinska Institutet & Stockholm Health Care Services, Stockholm, SE-11364, Sweden.
9. Department of Neurology, Alpert Medical School of Brown University, Providence, RI, 02903, USA.
10. Molecular & Cellular Therapeutics Department, Royal College of Surgeons in Ireland, Dublin, D15, Ireland.
11. Academic Unit Psychiatry of Old Age, University of Melbourne, Kew, VIC, 3101, Australia.
12. National Ageing Research Institute, Parkville, VIC, 3052, Australia.
13. Universite Lille, U1167 - RID-AGE - LabEx DISTALZ - Risk factors and molecular determinants of aging diseases, Lille, F-59000, France.
14. Institut National de la Sante et de la Recherche Medicale, U1167, Lille, F-59000, France.
15. Centre Hospitalier Universitaire de Lille, Department of Public Health, Lille, F-59000, Franch.
16. Institut Pasteur de Lille UMR1167, Lille, F-59000, France.
17. Division of Mental Health and Addiction, Oslo University Hospital, Oslo, 0407, Norway.
18. KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, 0407, Norway.
19. Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL, 60612, USA.
20. Department of Biomedical Engineering, Illinois Institute of Technology, Chicago, 60616, USA.
21. Departments of Psychiatry and Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands.
22. Department of Mathematics and Statistics, Curtin University, Perth, Australia.
23. CoE NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway, Oslo, 0455, Norway.
24. Centre for Clinical Brain Sciences and Edinburgh Imaging, University of Edinburgh, Edinburgh, EH16 4SB, United Kingdom.
25. Department of Biostatistics, School of Public Health, Boston University, Boston, MA, 02118, USA.
26. Framingham Heart Study, Chobanian and Avedisian Boston University School of Medicine,

Boston, MA, 02118, USA.
27. Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, 60612, USA.
28. Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, 98195-9458, USA.
29. Brain Center University Medical Center Utrecht, Utrecht, 3508GA, The Netherlands.
30. Vrije Universiteit, Amsterdam, 1081BT, The Netherlands.
31. Centre for Healthy Brain Ageing (CHeBA), Discipline of Psychiatry and Mental Health, School of Clinical Medicine, University of New South Wales, Sydney, NSW, 2052, Australia.
32. Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neurocience, VU Amsterdam, Amsterdam, 1081 HV, The Netherlands.
33. Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6525 EN, The Netherlands.
34. Institute of Clinical Chemistry and Laboratory Medicine, University Hospital Regensburg, Regensburg University, Regensburg, 93053, Germany.
35. LIFE Research Center for Civilization Diseases, University of Leipzig, Leipzig, 04103, Germany.
36. Department of Psychiatry, University Medical Center Utrecht, Utrecht, 3584CX, The Netherlands.
37. Altrecht Mental Health Institute, Utrecht, 3512PG, The Netherlands.
38. Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), {Georgia State, Georgia Tech, Emory}, Atlanta, GA, 30303, USA.
39. Pennington Biomedical Research Center, Baton Rouge, LA, 70808, USA.
40. Cerebral Imaging Centre, Douglas Research Centre, Montreal, QC, H4H 1R3, Canada.
41. Department of Psychiatry, McGill University, Montreal, QC, H3A 1A1, Canada.
42. Lieber Institute for Brain Development, Baltimore, MD, 21205, USA.
43. Imaging Genetics Center, Mark and Mary Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California, Marina del Rey, CA, 90292, USA.
44. Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, 52428, Germany.
45. Department of Biomedicine, University of Basel, Basel, CH-4031, Switzerland.
46. Medical Genetics, Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, 4031, Switzerland.
47. HU Virgen del Rocio, Instituto de Investigacion biomedica IBIS-CSIC, Universidad de Sevilla, CIBERSAM, Sevilla, 41013, Spain.
48. CNRS, IMN, UMR 5293, University of Bordeaux, Bordeaux, 33076, France.
49. Center for Multimodal Imaging and Genetics, La Jolla, CA, 92093, USA.
50. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, BS8 2BN, United Kingdom.
51. Population Health Sciences, University of Bristol, Bristol, BS8 BN, United Kingdom.
52. Department of Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, 1081 BT, The Netherlands.
53. Center for Translational & Computational Neuroimmunology, Department of Neurology, Columbia University Irving Medical Center, New York, NY, 10538, USA.
54. School of Psychology and Counselling, Queensland University of Technology, Brisbane, QLD, 4059, Australia.

55. INSERM U1219, Bordeaux Population Health Research Center, University of Bordeaux, Bordeaux, F-33000, France.
56. Department of Neurology, Institute of Neurodegenerative Diseases, Bordeaux University Hospital, Bordeaux, F-33000, France.
57. Imaging of Dementia and Aging Laboratory, Department of Neurology, University of California, Davis, Sacramento, CA, 95817, USA.
58. Department of Neurology, Hôpital Universitaire de Bruxelles, Université Libre de Bruxelles, Brussels, 1070, Belgium.
59. Social, Genetic, and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, SE5 8AF, United Kingdom.
60. Department of Medical Genetics, Oslo University Hospital, Oslo, 0450, Norway.
61. Department of Clinical Science, University of Bergen, Bergen, 5021, Norway.
62. Translational Developmental Neuroscience Section, Division of Psychological and Social Medicine and Developmental Neurosciences, Faculty of Medicine, TU Dresden, Dresden, 01307, Germany.
63. Department of Psychiatry and Psychotherapy, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, 11017, Germany.
64. Department of Psychology, University of Oslo, Oslo, 0373, Norway.
65. Department of Psychology, Oslo New University College, Oslo, 0456, Norway.
66. Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6500 HB, The Netherlands.
67. INSERM U1299, Paris Saclay University, Gif-sur-Yvette, 91190, France.
68. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, 6525 XD, The Netherlands.
69. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, 6500 HE, The Netherlands.
70. Department of Psychiatry and Behavioral Sciences, Rush University Medical Center, Chicago, IL, 60612, USA.
71. Department of Neurology, University of California Davis, Davis, CA, 95616, USA.
72. Institute of Molecular Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, TX, 77030, USA.
73. Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, 53127, Germany.
74. Department of Cognitive Neuroscience, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands.
75. Department of Human Genetics, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands.
76. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 02114, USA.
77. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, 17475, Germany.
78. Department of Medicine (Genetics), Mass General Brigham and Harvard Medical School, Boston, MA, 02115, USA.
79. Central Institute of Mental Health, Mannheim, 68159, Germany.
80. Goethe-University Frankfurt, Frankfurt am Main, 60528, Germany.

81. Department of Psychiatry and Mental Health, Neuroscience Institute, University of Cape Town, Cape Town, 7925, South Africa.
82. Section for Experimental Psychopathology and Neuroimaging, Department of General Psychiatry, Heidelberg University, Heidelberg, D-69115, Germany.
83. Icelandic Heart Association, Kopavogur, 201, Iceland.
84. Faculty of Medicine, University of Iceland, Reykjavik, 101, Iceland.
85. Department of Neuromedicine and Movement, NTNU Science, Trondheim, 7030, Norway.
86. MiDT National Research Center, Department of Research, St Olavs Hospital, Trondheim, 7006, Norway.
87. Norwegian Centre for Mental Health Research (NORMENT), Department of Mental Health and Addiction, University of Oslo, Oslo, 0450, Norway.
88. Centre for Forensic Psychiatry Research, Oslo University Hospital, Oslo, 0455, Norway.
89. Charité-Universitätsmedizin Berlin, Berlin, 10117, Germany.
90. Product Development, Genentech, Inc., South San Francisco, CA, 94080, USA.
91. Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, 117549, Singapore.
92. Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, University of Texas Health Sciences Center, San Antonio, TX, 78229-3900, USA.
93. Department of Population Health Sciences, UT Health Science Center San Antonio, San Antonio, TX, 78229, USA.
94. Department of Neurology, Boston University School of Medicine, Boston, MA, 02118, USA.
95. Department of Psychiatry, Carver College of Medicine, University of Iowa, Iowa City, IA, 52246, USA.
96. Max Planck Institute of Psychiatry, Munich, 80804, Germany.
97. Department of Child and Adolescent Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, 9713 GZ, The Netherlands.
98. Accare Child Study Center, Groningen, 9723 HE, The Netherlands.
99. Division of Neurogeriatrics, Department of Neurology, Medical University of Graz, Graz, 8036, Austria.
100. Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, 8036, Austria.
101. German Centre for Neurodegenerative Diseases (DZNE) – site Rostock/Greifswald, Greifswald, 17489, Germany.
102. Institute for Community Medicine, University Medicine Greifswald, Greifswald, 17495, Germany.
103. Department of Psychiatry, Brain Health Institute, Rutgers University, Piscataway, NJ, 08854, USA.
104. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, 17475, Germany.
105. Department of Radiology, University Clinic Greifswald, Greifswald, 17475, Germany.
106. Departments of Epidemiology and Neurology, Erasmus MC, Rotterdam, 3015 CN , The Netherlands.
107. Department of Psychiatry and Mental Health, Neuroscience Institute, Groote Schuur Hospital, University of Cape Town, Cape Town, 7925, South Africa.
108. Department of Radiology, Mayo Clinic, Rochester, MN, 55905, USA.
109. Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet

& Stockholm Health Care Sciences, Stockholm Region, Stockholm, SE-11364, Sweden.
110. Araya, Inc., Tokyo, 101-0025, Japan.
111. Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, 3015 GD, The Netherlands.
112. Intramural Research Program, National Institute on Aging, Baltimore, MD, 21224, USA.
113. Division of Psychiatry, University of Edinburgh, Edinburgh, EH10 5HF, United Kingdom.
114. Center for Genomic Medicine, Mass General Brigham, Boston, MA, 02114, USA.
115. Department of Psychiatry, Harvard Medical School, Boston, MA, 02115, USA.
116. Stanley Center for Psychiatry, Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.
117. Groupe d'Imagerie Neurofonctionnelle, Institut des Maladies Neurodégénératives, UMR 5293, CNRS, Université de Bordeaux, Bordeaux, 33076, France.
118. PPLS, University of Edinburgh, Edinburgh, EH8, United Kingdom.
119. Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, 01655, USA.
120. Department of Neurology, University of Washington, Seattle, WA, 98104-2420, USA.
121. Department of Epidemiology, University of Washington, Seattle, WA, 98195-9458, USA.
122. Departments of Neurology and Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, 15213, USA.
123. Department of Psychology, University of Edinburgh, Edinburgh, EH8 9JZ, United Kingdom.
124. Université Paris-Saclay; Institut National de la Santé et de la Recherche Médicale, INSERM U1299 "Trajectoires développementales Psychiatrie", Ecole Normale Supérieure Paris-Saclay, CNRS UMR 9010, Université Paris Cité, Centre Borelli, Gif sur Yvette, 911.
125. School of Clinical Sciences, Queensland University of Technology, Brisbane, QLD, 4001, Australia.
126. Institute of Gerontology and Geriatrics, Department of Medicine and Surgery, University of Perugia, Perugia, 06132, Italy.
127. Clinical Geriatrics, NVS Department, Karolinska Institute, Huddinge, 14152, Sweden.
128. Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, 68159, Germany.
129. Institute of Neurogenomics,Helmholtz Munich, 85764, Neuherberg, Germany.
130. Neurogenetic Systems Analysis Group, Institute of Neurogenomics, Helmholtz Munich, 85764, Neuherberg, Germany.
131. Department of Psychiatry, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, 1081 HJ, The Netherlands.
132. Amsterdam Public Health, Mental Health program, Amsterdam, 1081 BT, The Netherlands.
133. Amsterdam Neuroscience, Mood, Anxiety, Psychosis, Sleep & Stress program, Amsterdam, 1081 BT, The Netherlands.
134. Amsterdam Neuroscience, Complex Trait Genetics program, Amsterdam, 1081 HV, The Netherlands.
135. MIND Center, Jackson, MS, 39216, USA.
136. Cécile and Oskar Vogt Institute for Brain Research, Medical Faculty, University Hospital Düsseldorf, Heinrich Heine University Düsseldorf, Düsseldorf, D-40225, Germany.
137. Department of Biomedicine, University Hospital Basel and University of Basel, Basel, CH-4031, Switzerland.
138. Statistics Genetics Group, Max Planck Institute of Psychiatry, Munich, 80804, Germany.

139. Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, 17489, Germany.
140. German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, 17489, Germany.
141. Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, 46202, USA.
142. University Medical Center Groningen, Groningen, 9713GZ, The Netherlands.
143. Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD, 21287, USA.
144. General internal Medicine, Johns Hopkins School of Medicine, Baltimore, MD, 21287, USA.
145. Clinical Neuropsychology section, Vrije Universiteit Amsterdam, Amsterdam, 1081 BT, The Netherlands.
146. Emma Children's Hospital, University Medical Centers Amsterdam, Amsterdam, 1100 DD, The Netherlands.
147. Amsterdam Reproduction & Development Research Institute, Amsterdam, 1100 DD, The Netherlands.
148. Université Libre de Bruxelles, Brussels, 1070, Belgium.
149. Department of Neurology and Neurosurgery, McGill University, Montreal, QC, H3A 2B4, Canada.
150. Departments of Psychiatry and Neuroscience, Faculty of Medicine, University of Montreal, Montreal, QC, H3T 1C5, Canada.
151. Centre Hospitalier Universitaire Sainte-Justine, University of Montreal, Montreal, QC, H3T 1C5, Canada.
152. Hospital for Sick Children, Toronto, ON, M5G 0A4, Canada.
153. Department of Physiology, University of Toronto, Toronto, M5G 0A4, Canada.
154. Departments of Radiology and Clinical Neurosciences, Hotchkiss Brain Institute, Cumming School of Medicine, University of Calgary, Calgary, AB, T2N 4N1, Canada.
155. Department of Health Systems and Population Health, Seattle, WA, 98195-9458, USA.
156. Translational Psychiatry, Munich, 80804, Germany.
157. Department of Developmental Disability Neuropsychiatry, Discipline of Psychiatry and Mental Health, School of Clinical Medicine, University of New South Wales, Sydney, NSW, 2052, Australia.
158. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Faculty of Medicine Mannheim, University of Heidelberg, Mannheim, 68159, Germany.
159. Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, 46202, USA.
160. Indiana Alzheimer's Disease Research Center, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, 46202, USA.
161. Department of Psychiatry and Neuroscience, Charité - Universitätsmedizin Berlin, Berlin, 10117, Germany.
162. Department of Psychology, Clinical Psychology and Psychotherapy, MSB Medical School Berlin, Berlin, 14197, Germany.
163. Instituto de Biomedicina de Sevilla (IBiS) HUVR/CSIC/Universidad de Sevilla/ CIBERSAM, ISCIII, Dpto. de Fisiología Médica y Biofísica, Sevilla, 41013, Spain.
164. Department of Psychiatry, University of Cambridge, Cambridge, CB2 0SZ, United Kingdom.
165. Department of Radiology and Nuclear Medicine, Erasmus MC University Medical Center, Rotterdam, 3015 GD, The Netherlands.

166. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA.
167. Neuropsychiatric Institute, The Prince of Wales Hospital, Randwick, NSW, 2031, Australia.
168. Centre for Youth Mental Health, The University of Melbourne, Parkville, VIC, 3052, Australia.
169. Orygen, Parkville, VIC, 3052, Australia.
170. Institute of Molecular Biology & Biochemistry, Gottfried Schatz Center for Signaling, Metabolism & Aging, Medical University Graz, Graz, 8010, Austria.
171. Department of Neurology, Medical University Graz Austria, Graz, 8023, Austria.
172. Neuroscience Research Australia, Sydney, NSW, 2031, Australia.
173. School of Biomedical Sciences, University of New South Wales, Sydney, NSW, 2052, Australia.
174. Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, 04107, Germany.
175. Centre for Population Neuroscience and Stratified Medicine (PONS), ISTBI, Fudan University, Shanghai, 200031, P.R. China.
176. PONS Centre, Department of Psychiatry, CCM, Charite Unversitaetsmedizin Berlin, Berlin, 10017, Germany.
177. Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA.
178. The Hospital for Sick Children, Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto, ON, M5G 0A4, Canada.
179. Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology, London, WC1N 3BG, United Kingdom.
180. Chalfont Centre for Epilepsy, Chalfont St Peter, SL9 0RJ, United Kingdom.
181. Department of Biostatistics, University of Michigan, Ann Arbor, MI, 48109, USA.
182. Department of Neurology, Institute of Clinical Mediciine, University of Eastern Finland, Kuopio, 70100, Finland.
183. Dr. Einar Martens Research Group for Biological Psychiatry, Department of Medical Genetics, Haukeland University Hospital, Bergen, N-5021, Norway.
184. SAMRC Research Unit on Risk & Resilience in Mental Disorders, Department of Psychiatry & Neuroscience Institute, University of Cape Town, Cape Town, 7925, South Africa.
185. Department of Genetics & UNC Neuroscience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599-7250, USA.
186. Instituto de Física de Cantabria (CSIC-UC), Santander, E-39005, Spain.
187. Department of Radiology, Marqués de Valdecilla University Hospital, Valdecilla Biomedical Research Institute IDIVAL, Santander, 39011, Spain.
188. The National Centre of Excellence in Intellectual Disability Health,, Faculty of Medicine and Health, University of New South Wales, Sydney, NSW, 2052, Australia.
189. Department of Biological Psychology & Netherlands Twin Register, Vrije Universiteit Amsterdam, Amsterdam, 1081 BT, The Netherlands.
190. Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, 6525 GA, The Netherlands.
191. School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, 6200MD, The Netherlands.

192. Department of Psychiatry, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands.
193. Department of Psychiatry, University Hospital Marqués de Valdecilla - IDIVAL, Santander, 39008, Spain.
194. Departamento de Medicina y Psiquiatría, Universidad de Cantabria, Santander, 39008, Spain.
195. Centro de Investigación Biomédica en Red en Salud Mental (CIBERSAM), Sevilla, 41013, Spain.
196. Department of Neurology, Max Planck Institute for Human, Cognitive and Brain Sciences, Leipzig, 04103, Germany.
197. Cognitive Neurology, University of Leipzig Medical Center, Leipzig, 04103, Germany.
198. Department of Psychiatry, Massachusetts General Hospital, Boston, MA, 02114, USA.
199. UK Dementia Research Institute Centre, University of Edinburgh, Edinburgh, EH16 4SB, United Kingdom.
200. Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA.
201. Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA.
202. Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, 21205, USA.
203. University of California San Francisco, San Francisco, CA, 94121, USA.
204. Northern California Institute for Research & Education (NCIRE), San Francisco, CA, 94121, USA.
205. Veterans Administration Medical Center, San Francisco, CA, 94121, USA.
206. Division of Clinical Geriatrics, Department of Neurobiology, Care Sciences and Society (NVS), Karolinska Institutet, Huddinge, 14183, Sweden.
207. Section on Social and Cognitive Developmental Neuroscience, National Institute of Mental Health, Bethesda, MD, 20892-1276, USA.
208. Department of Population Health Sciences and Glenn Biggs Institute for Alzheimer's and Neurodegenerative Diseases, UT Health San Antonio, San Antonio, TX, 78229, USA.
209. Department of Neurology, Boston University Chobanian & Avedisian School of Medicine, Boston, MA, 02118, USA.
210. School of Psychology, The University of Queensland, Brisbane, QLD, 4072, Australia.

* These authors contributed equally to this work.

Corresponding authors: Luis M. García-Marín (luis.garciamarin@qimrberghofer.edu.au) and Miguel E. Rentería (miguel.renteria@qimrberghofer.edu.au).

Abstract

Subcortical brain structures are involved in developmental, psychiatric and neurological disorders. We performed GWAS meta-analyses of intracranial and nine subcortical brain volumes (brainstem, caudate nucleus, putamen, hippocampus, globus pallidus, thalamus, nucleus accumbens, amygdala and, for the first time, the ventral diencephalon) in 74,898 participants of European ancestry. We identified 254 independent loci associated with these brain volumes, explaining up to 35% of phenotypic variance. We observed gene expression in specific neural cell types across differentiation time points, including genes involved in intracellular signalling and brain ageing-related processes. Polygenic scores for brain volumes showed predictive ability when applied to individuals of diverse ancestries. We observed causal genetic effects of brain volumes with Parkinson's disease and ADHD. Findings implicate specific gene expression patterns in brain development and genetic variants in comorbid neuropsychiatric disorders, which could point to a brain substrate and region of action for risk genes implicated in brain diseases.

Introduction

Subcortical brain structures are affected in most major neurological diseases, including psychiatric and developmental brain disorders[1]. These brain structures are involved in crucial daily functions, such as learning[2,3], memory[3,4], attention[3], motor control[2,3], and reward[5,6]. Likewise, intracranial volume (ICV) variation has been associated with neuropsychiatric phenotypes in observational[7,8] and genetic[9–11] studies. Notably, genome-wide association studies (GWAS) have revealed a shared genetic aetiology between brain structures and behavioural, neuropsychiatric, and other health-related phenotypes[2,12–15].

While neuroimaging genetic studies have advanced our understanding of the genetic architecture of subcortical[2,16] and cortical[13,17] brain structures, the most highly powered studies have uncovered the genetic underpinnings of the global measures of the cortex and specific cortical brain structures[13,18,19]. Therefore, there is a need to leverage large and diverse datasets to uncover genetic variants that provide insights into the mechanistic pathways responsible for variation in the volumes of intracranial and subcortical brain volumes.

We coordinated a worldwide analysis of 49 study samples from 19 countries and conducted the largest international genetic analysis of human subcortical brain volumes and ICV. We analysed individual and summary-level genetic data from participants across four international sources to accomplish three goals. First, we sought to characterise the genetic and molecular underpinnings of intracranial and nine subcortical brain volumes (i.e. the brainstem, caudate nucleus, putamen, hippocampus, globus pallidus, thalamus, nucleus accumbens, amygdala and, for the first time, the ventral diencephalon). We performed GWAS meta-analyses including over 70,000 individuals, investigated the genetic overlap among these structural brain volumes, and conducted gene-based tests, eQTL mapping with transcriptome-wide association studies, and the integration of single-cell RNA sequencing data with GWAS summary statistics. Second, we evaluated the predictive utility of polygenic scores for these brain volumes in a diverse ancestral population. Finally, we investigated the overlap and potential causal genetic effects between the observed brain-associated genomic loci and genomic markers implicated in major neurological and psychiatric diseases to examine structure-specific genetic associations with major brain diseases.

This work is crucial, as it can point to a brain substrate and region of action for risk genes implicated in brain diseases.

Results

Genome-wide association analyses

We identified 529 genome-wide significant loci (*p*-value < $5x10^{-8}$) associated with human intracranial or subcortical brain volumes (**Table 1 and Supplementary Figures 1 - 20**), of which 367 survived a multiple testing correction for the total number of phenotypes (*p*-value < $6.25x10^{-9}$). Of the 529 genome-wide significant loci (**Supplementary Tables 1 and 2**), 254 were independent unique loci across structures (**Supplementary Table 3**). Brainstem volume showed the largest number of independent genetic associations, whereas the amygdala volume had the fewest (**Figure 1** and **Table 1**). SNP-based heritability estimates indicated that common genetic variants explained a substantial proportion of the phenotypic variation of intracranial and subcortical brain volumes, ranging from 17% for the volume of the amygdala to 35% for the volume of the brainstem (**Table 1**). LD score regression intercepts close to or equal to 1 suggested that the elevated lambdas and inflation in the quantile plots (**Supplementary Figures 1 - 20**) were most likely due to polygenicity rather than population stratification (**Table 1**). Attenuation ratios close to 0 indicated correct genomic control. Manhattan and QQ plots for GWAS in individual cohorts are available in **Supplementary Figures 21 - 60**.

As a sensitivity analysis, we performed GWAS in the UK Biobank cohort for subcortical brain volumes without adjusting for ICV (**see Methods and Supplementary Figures 61 - 78**.). Direction and magnitude of SNP effect sizes were largely consistent as suggested by Pearson's correlations using the SNP effect sizes for the same phenotype with and without the adjusting for ICV (correlations range = 0.81 - 0.92). Moreover, we split the UK Biobank sample into two randomised subsamples (N ~ 18,047 each) in an attempt to investigate replicability for intracranial and subcortical brain volumes (**Supplementary Figures 79 - 118)**. Direction and magnitude of SNP effect sizes were for the most part consistent as suggested by Pearson's correlations using the effect sizes for the same phenotype for both subsamples (correlations range = 0.67 - 0.84).

Functional annotation and gene prioritisation

We used MAGMA (v1.08) to perform gene-based association analyses. GWAS meta-analysis for ICV and the volumes of the brainstem and caudate nucleus showed the largest number of genes associated with each structure, followed by the volumes of the putamen, hippocampus, ventral diencephalon, globus pallidus, thalamus, and nucleus accumbens (**Supplementary Table 4**). Amygdala volume was associated with the fewest genes. No single gene was associated with all intracranial or subcortical brain volumes, which reflects the correction for ICV. The Forkhead Box O3 (*FOXO3*) gene was associated with the volume of five brain structures. Similarly, the Geminin Coiled-Coil Domain Containing (*GMNC*), A-kinase anchoring protein 10 (*AKAP10*), Epidermal Growth Factor Receptor (*EGFR*), Microtubule Nucleation Factor (*TPX2*), and the Bcl-2-like Protein 1 (*BCL2L1*) were associated with the volume of four brain structures. Furthermore, genes from the *HOX* and *PAX* homeobox gene families were associated with the volume of the brainstem. In addition, genes from the *WNT* family were associated with brainstem, ventral diencephalon and intracranial volumes. Other genes associated with multiple subcortical brain volumes included *BIRC6, CRHR1, IGF1, MAPT, NUP37, NUP43, KTN1, FOXS1* and *COX4I2*, which have been previously reported to have roles in intracellular signalling[20], autophagy[21–23], and multiple brain ageing processes, such as vascular ageing, oxidative resistance, tau pathology, and apoptosis[24–27]. A full list of statistically significant gene-based test findings after Bonferroni multiple testing correction is available in **Supplementary Table 5.**

We integrated our GWAS results with expression quantitative trait loci (eQTL) data from the Genotype-Tissue Expression project (GTEx/v8) (**Supplementary Table 6**). We observed consistent findings with our gene-based tests. The genes *CRHR1, NUP43,* and *KTN1* were associated with subcortical brain volumes. Furthermore, we observed associations for the genes *UQCC1,* and *COX4I2.* Genes that may be linked to specific brain structures through changes in gene expression include, among others, *CRHR1* for the putamen, *FAIM* for the thalamus, and *MAPK3* as well as *ZNF786* for the hippocampus. We prioritised potential causal genes from the associated loci performing transcriptome-wide association studies (TWAS). Most genes were associated uniquely with the volume of a single brain structure (91%), while others were shared across the volumes of several brain structures. With this approach we observed associations of the genes *CRHR1, MAPT, NUP43, NUDT14, FAIM, MAPK3, and ZNF786* with subcortical brain volumes (**Supplementary Table 6**), even after correcting for multiple testing using a conservative

approach (*p*-value < 3.06×10^{-4}, **see Methods**). Likewise, we revised eQTLs in developmental datasets to identify genes involved in brain development (**Supplementary Table 7**) and observed associations of brainstem, caudate nucleus, putamen, thalamus, ventral diencephalon, and intracranial volumes with the genes *LRRC37A, LRRC37A2, KANSL1, RPS26, ARL17B, PILRB, PILRA*, and *EFCAB13* after correcting for multiple testing using a conservative approach (*p*-value < 1.26×10^{-3}, **see Methods**).

We integrated single-cell RNA-sequencing data[28] with GWAS summary statistics to identify critical cell types and cellular processes influencing intracranial and subcortical brain volumes variation. From the prioritised genes across MAGMA and TWAS analyses, we identified nine expressed genes (*TUFM, CRHR1, NUP43, MAPK3, LRRC37A2, FAIM, ZNF786, YIPF4* and *PSMC3*) across seven different cell types, including pluripotent floor progenitor plate cells (FPP), proliferating floor progenitor plate cells (P_FPP), dopaminergic neurons (DA), ependymal-like 1 (Epen1), serotonergic-like neurons (Serts), and astrocyte-like cells (Astro), influencing brain volume variation. Our gene expression findings in cell types mentioned previously cover up to 52 days of differentiation. Most of the expressed genes at day 11 were observed in FPP and P_FPP; at day 30 in FPP, DA, and ependymal-like 1 cells; and at day 52 in DA, serotonergic-like neurons, astrocyte-like, ependymal-like 1 cells. Full results surviving multiple testing correction (p-value < 1.19×10^{-3}, **see Methods**) are available in **Supplementary Table 8**.

As a sensitivity analysis, we performed MAGMA analyses for subcortical brain volumes using data from the UK Biobank with and without adjusting for ICV (**Supplementary Tables 9 and 10**). Identified genes were consistent with and without the adjustment for ICV. However, these genes were associated with more subcortical brain volumes when GWAS were not adjusted for ICV.

Polygenic scores predict phenotypic brain volumes

We tested the predictive capability of our genome-wide results by performing the meta-analyses leaving out the ABCD cohort (N = 5,267) to determine whether polygenic scores from European ancestry samples are associated with intracranial and subcortical brain volumes in the more diverse ABCD cohort. The polygenic scores for all brain volumes were strongly associated with intracranial and subcortical volumes in participants of European, African, and non-European

ancestries, as well as across all ancestral groups (**Figure 2** and **Supplementary Figures 119-124**). Overall, results remained consistent with additional adjustments for cryptic relatedness. While polygenic prediction was most accurate for participants of European ancestry (variance explained ranging from 2.1 to 8.5%), we observed that the variance explained in non-European ancestry groups was also significant and ranged from 0.8 to 9.8% (**Figure 2** and **Supplementary Table 11**). Sensitivity analyses included linear regressions among participants of European ancestry for subcortical volumes using ICV as a covariate. The results were consistent and remained essentially unchanged. As expected, polygenic scores for ICV did not explain residual variance above phenotypic ICV (**Supplementary Table 12, Supplementary Table 13,** and **Supplementary Figure 125**).

Genetic overlap between subcortical brain structures

Using LD score regression, we estimated genetic correlations among intracranial and the nine subcortical brain volumes under study. We adopted a conservative approach to multiple testing and corrected for the total number of genetic correlations, including those for other complex human phenotypes [0.05 / 320 [total number of genetic correlation tests] = 1.56×10^{-4}]. We observed substantial genetic overlap among intracranial and subcortical brain volumes (**Figure 3** and **Supplementary Tables 14 - 15**). The thalamus volume showed genetic correlations with the other eight brain volumes. The volume of the brainstem, amygdala, and the caudate nucleus, with four significant genetic correlations, showed the fewest. Components of the striatum, including the caudate nucleus and putamen, were strongly correlated with the nucleus accumbens**.** Within-phenotype genetic correlations across cohorts were large (rg > 0.60) and statistically significant after multiple testing correction (**Supplementary Table 16**).

We further explored polygenic overlap between the GWAS summary statistics, including all cohorts, for the subcortical brain volumes using MiXeR. We estimated the number of causal variants influencing each subcortical brain volume (median n causal variants = 1.92K, **Supplementary Table 17**). The volume of the hippocampus was the least polygenic (1.000K causal variants, SE = 0.13K), whilst thalamus volume was the most polygenic (2.58K causal variants, SE = 0.15K). We then estimated the number of causal variants shared between subcortical brain volumes, finding substantial polygenic overlap between them (median n shared

causal variants = 1.24K, **Supplementary Table 18**). The largest overlap was observed between the volumes of the thalamus and globus pallidus (2.08K variants, SE = 0.22K), while the smallest was between the thalamus and hippocampus (0.53K variants, SE = 0.04K). We identified polygenic overlap between three pairs of brain structures: brainstem-amygdala (0.97K variants, SE = 0.09K), brainstem-caudate nucleus (0.87K variants, SE = 0.10K), and caudate-ventral diencephalon (1.01K variants, SE = 0.12K), despite their genetic correlation being close to zero (**Figure 3**).

Genetic clustering of subcortical brain structures

We used genomic structural equation modelling (SEM)[29] to examine whether and how subcortical brain structures cluster together at a genetic level. We first tested a common factor model, which provided a poor fit to the data (CFI = 0.70, SRMR = 0.13, AIC = 828.03; **Supplementary Table 19**). To explore other possible factor structures underlying subcortical brain structures, we conducted genetic exploratory factor analyses (EFA) based on the genetic correlation matrix of the 9 subcortical structures (**Supplementary Table 20**). A two-factor model (**Supplementary Table 21**) and three-factor model (**Supplementary Table 22**) explained 43% and 53% of the total genetic variance, respectively. Follow up confirmatory factor analyses (CFA) were specified in genomic SEM (retaining standardised loadings greater than 0.25). While the two-factor model did not provide adequate fit (CFI = 0.84, SRMR = 0.09, AIC = 482.97), the three-factor model provided good fit to the data (CFI = 0.91, SRMR = 0.06, AIC = 299.33, **Figure 4**).

Genetic correlations with brain disorders

We estimated genetic correlations between the brain volumes investigated and 22 complex human phenotypes (**Figure 3** and **Supplementary Tables 14 - 15**). Parkinson's disease, attention-deficit/hyperactivity disorder (ADHD), neuroticism score, birth weight, birth head circumference, height, and insomnia showed statistically significant associations after correction for multiple testing. Parkinson's disease showed several positive genetic correlations with intracranial and subcortical brain volumes, including those of the nucleus accumbens, brainstem, caudate nucleus, globus pallidus, putamen, thalamus, and ventral diencephalon. We observed negative genetic overlap for ICV with ADHD, insomnia, and neuroticism scores. Conversely, we

identified a positive genetic correlation of birth weight, birth head circumference, and height with ICV.

We further investigated the relationship between brain volumes and complex human phenotypes with a statistically significant genetic correlation using the pairwise-GWAS (GWAS-PW) method. With this approach, we identified 338 genomic segments with genetic variants influencing both the volume of a brain structure and a human complex phenotype (**Supplementary Table 23**). Genomic segments with shared genetic variants were identified for all traits that displayed a significant genetic correlation after multiple testing correction, except for the ventral diencephalon and ADHD.

As a sensitivity analysis, we investigated whether adjusting or not adjusting for ICV had an effect on genetic correlations with complex human phenotypes. We used the GWAS for subcortical brain volumes from the UK Biobank with and without adjusting for ICV and estimated genetic correlations with complex human traits. We observed more statistically significant genetic correlations with complex human phenotypes when not adjusting subcortical brain volumes for ICV. However, the direction and magnitude of the genetic correlations remained for the most part consistent regardless of the adjustment for ICV (**Supplementary Tables 24 - 29; Supplementary Figures 126 and 127**).

Potential causal genetic effects

We estimated the genetic causal proportion (GCP) with the latent causal variable method (LCV) and leveraged the Latent Heritable Confounder Mendelian Randomisation (LHC-MR) method to assess potential causal genetic effects of intracranial and subcortical brain volumes with complex human traits that displayed a statistically significant genetic correlation after Bonferroni multiple testing correction. We observed putative causal genetic effects for a larger putamen volume influencing a higher risk for Parkinson's disease after multiple testing correction using the LCV [0.05 / 16 [total number of genetic causal proportion tests in the present study] = 3.13×10^{-3}] and LHC-MR [0.05 / 32 [total number of LHC-MR tests in the present study] = 1.56×10^{-3}] methods. With both methods, we observed that a larger ICV could reduce the likelihood of developing ADHD. Potential causal genetic effects suggesting that a larger ICV could reduce the likelihood of

developing insomnia were observed with LCV, but not with LHC-MR. We observed several potential causal genetics effects of nominal significance (p-value < 0.05), which are fully described in **Supplementary Table 30** and **Supplementary Table 31**

Discussion

We performed the largest GWAS meta-analysis of intracranial and subcortical brain volumes to date across international datasets from 19 countries. Here, we complement and extend work from a previous GWAS meta-analysis that identified 48 significantly associated loci with seven subcortical brain volumes[2]. Our results implicated more than 254 independent genetic variants, at the common genome-wide threshold (*p-value* < $5x10^{-8}$), associated with ICV or the volumes of the brainstem, caudate nucleus, putamen, hippocampus, globus pallidus, thalamus, nucleus accumbens, amygdala and for the first time, the ventral diencephalon, in over 70,000 individuals. Of these 254 independent genetic variants, 161 have not been reported in previous studies[2,14–16,30]. From the independent genome-wide genetic variants reported in previous studies[2,14–16,30], we replicated 39% (N = 167) in the genome-wide loci in our meta-analysis at the common genome-wide threshold (*p*-value < $5x10^{-8}$). Our findings provide insights into genes that influence variation in human intracranial and subcortical brain volumetric measures. We show that distinct genetic variants often have a specific effect on the variation of a single brain volume. In addition, we conducted thorough functional annotation and gene prioritisation analyses, including gene-based tests, TWAS, and the integration of single-cell RNA sequencing data with GWAS summary statistics. We investigated the genetic overlap and putative causal genetic effects of intracranial and subcortical brain volumes with other complex human phenotypes. Polygenic scores for intracranial and subcortical volumes showed predictive ability for their corresponding phenotypic measurements, even when examined in a pre-adolescent population with individuals of diverse ancestral backgrounds.

Previous work suggests that heritability estimates for intracranial and subcortical brain volumes range from 33 to 86% in twin and family studies[2,31,32] and from 9 to 33% using a SNP-based heritability approach[2]. In our study, SNP-based heritability estimates derived from GWAS meta-analysis results ranged from 18 to 38%. These values are consistent with prior findings in the UK Biobank and ENIGMA cohorts[2,33]. Furthermore, a previous GWAS meta-analysis of ICV identified 64 genetic variants explaining 5% of phenotypic variation in a sample of European ancestry[14]. In the present study, we explained 28% (CI = 26 - 30%) of phenotypic variation and

identified 83 significant loci associated with ICV at the common genome-wide threshold (*p*-value < $5x10^{-8}$).

We explored genetic correlations among intracranial and subcortical brain volumes, including the first ever findings for the ventral diencephalon. We identified substantial genetic overlap for these brain volumes, consistent with previous reports, and supporting previously observed phenotypic associations[2]. In contrast with previous findings[2], we identified several genetic correlations of subcortical brain volumes, including the hippocampus, globus pallidus, thalamus and ventral diencephalon, with the brainstem. The strongest genetic correlation among all brain volumes was observed between the volumes of brainstem and the ventral diencephalon. This finding is consistent with brainstem anatomy and the interconnection with the ventral diencephalon, as the brainstem can be subdivided into the diencephalon (thalamus, hypothalamus), mesencephalon (midbrain), ventral metencephalon (pons), and myelencephalon (medulla)[34]. In addition, with genomic SEM analyses we observed how subcortical brain structures cluster together at a genetic level. The volumes of the nucleus accumbens, caudate nucleus, putamen and globus pallidus clustered together, which is consistent with the structure of the basal ganglia[35–37] and the striatum[36]. Furthermore, the volume of the globus pallidus also clustered together with those of structures strongly interconnected with the basal ganglia[35], such as the brainstem, thalamus, and ventral diencephalon, while the volumes of the amygdala and the hippocampus, whose circuitry in the limbic system is well-known to predominantly influence emotion-regulated memories[38], constituted the third cluster.

Previous studies have aimed to investigate the genetic overlap of intracranial and subcortical brain volumes with neuropsychiatric disorders[1,2,12,39,40]. Here, we identified genetic correlations for eight subcortical brain volumes with Parkinson's disease and three with ADHD. ICV showed genetic overlap with both Parkinson's disease and ADHD. ADHD and Parkinson's disease are predominantly young- and late-onset phenotypes, respectively[41,42]. However, our GWAS summary statistics for intracranial and subcortical brain volumes do not necessarily include people diagnosed with Parkinson's disease, ADHD, or individuals at high risk for these disorders. Thus, positive genetic correlations with Parkinson's disease suggest that genetic variants influencing larger volumes during the development of specific structures are also associated with a higher

risk for Parkinson's disease, consistent with previous observations in genetic studies[2]. In contrast, negative genetic correlations with ADHD imply that genetic variants influencing a smaller volume of specific structures are associated with a higher genetic susceptibility for ADHD[43]. We present, for the first time, the further interrogation of the observed genetic correlations via different methods to demonstrate putative causal genetic effects between a range of subcortical brain volumes and various complex human phenotypes.

Identified loci for intracranial and subcortical brain volumes were annotated using gene-based testing, eQTL mapping, TWAS and the integration of single-cell RNA sequencing data with GWAS summary statistics. Most of the genes associated with intracranial or subcortical brain volumes across analyses were uniquely associated with a specific brain volume, shedding light on the independent genetic underpinnings of these structures. While the remaining genes showed effects influencing more than one brain structure, no single gene was associated with all brain measures assessed. We identified gene expression in different neural cell types for genes that have been previously reported to act through pathways related to autophagy (*TUFM,* and *FAIM*)[21–23], mediation of intracellular signalling (*MAPK3*)[20], organelle biogenesis and maintenance (*YIPF4*)[44], and nucleo-cytoplasmic transport of RNA and proteins (*NUP43*)[45]. Some of the identified expressed genes (*CRHR1*[24–26] and *LRRC37A2*[46,47]) have been previously associated with neurodegenerative disorders[48]. For instance, it has been suggested that *CRHR1* may have a neuroprotective effect in Parkinson's disease[24–26], and may even prevent dementia-related symptoms[49].

Previous studies suggest that polygenic scores lack predictive ability on ancestral groups that do not match the ancestry of the discovery GWAS[50]. However, in the present study, we observed for the first time that polygenic scores significantly predicted the same intracranial and subcortical brain volumes in a sample of preadolescent children of European and non-European ancestries. Given that PRS prediction was possible in children, it is likely that the genetic variation underlying differences in adult intracranial and subcortical brain volumes is present at an early age. This is consistent with prior work suggesting that prenatal and postnatal development of subcortical brain regions is influenced by genetic variants associated with subcortical brain volumes in adults[51]. Furthermore, our polygenic scores account for a significant fraction of brain

variability across ancestries. This suggests that genetic variants responsible for subcortical brain structure could be shared across ancestries, with linkage disequilibrium and minor allele frequency differences underlying differences in accuracy for trans-ancestry predictions[52]. We observed that predictions on participants of African ancestry outperformed those for participants of Asian ancestry. This is inconsistent with previous studies demonstrating that LD patterns in Asians are more similar to those in Europeans when compared with those of African ancestry. We attribute our observations to the difference in sample sizes, which is larger for participants of African ancestry (N = 1,833) than for those of Asian ancestry (N = 152). Overall, our findings point towards polygenic score generalisability across individuals of diverse ancestral backgrounds, and could be leveraged to study brain development in young populations. Well powered polygenic predictors, will potentially enable to boost power of future neuroimaging GWAS performed in samples of underrepresented ancestries[53], an important endeavour to narrow the ancestry biases in current genetic studies.

When performing GWAS on any brain measurements, the inclusion of ICV as a covariate in the model is frequently used and widely accepted to adjust for differences in head size among participants[2,13,16,30]. However, this practice remains open for discussion as there is potential for collider bias. Correcting for a heritable, correlated covariate, such as ICV, can bias estimates, which could potentially limit the interpretability of gene-identification and other downstream analyses[54]. In the present study, we performed GWAS for subcortical brain volumes in the UK Biobank cohort with and without adjusting for ICV to investigate potential differences. We estimated genetic correlations with complex human phenotypes and performed gene-based tests. For these analyses, we observed more statistically significant associations for the GWAS that were not adjusted for ICV. We suggest that the effect of ICV is driving these associations. For instance, ICV is correlated with head birth circumference and birth weight. When not adjusting for ICV, most if not all of the subcortical brain volumes were genetically correlated with head birth circumference and birth weight after multiple testing correction. Consistently, when adjusting for ICV, a few subcortical brain volumes were barely genetically correlated with these phenotypes. Similar observations were made for gene-based tests. Lastly, when a correction for ICV is not included in volumetric studies using magnetic resonance imaging, sex differences are observed[55]. We consider that for the analyses of brain size-related measurements, the adjustment for ICV is

necessary to account for differences in head size and sex, which will directly influence the measurements. We consider this crucial in our study since we leveraged data from different cohorts, such as ABCD and the UK Biobank, which include participants of different age, sex, and total brain size. Future studies should aim to fully investigate the effect of ICV on neurogenomic analyses.

The limitations of this study must be acknowledged. As we mentioned in the methods section, the imaging analysis and visualisation of structural data in all cohorts was performed using the publicly available FreeSurfer package tool, which includes the superior cerebellar peduncle as part of the brainstem. The superior cerebellar peduncle is a structure that connects the cerebellum to the brainstem[56]. However, anatomically, the cerebellar peduncle is not a putative structure of the brainstem[56]. Therefore, we note that the inclusion of the cerebellar peduncle as part of the volume of the brainstem is a limitation of the segmentation performed by the FreeSurfer package tool, which we are unable to address. Furthermore, our GWAS meta-analyses included only participants of European ancestry in the discovery phase. Therefore, the genetic loci associated with intracranial and subcortical brain volumes in the present study are only representative of individuals of European ancestry until confirmed in samples of other ancestral populations.

We provide evidence for the polygenic architecture of intracranial and subcortical brain volumes, presenting findings for the volume of the ventral diencephalon for the first time, and show that polygenic scores could be useful in predicting or imputing brain volume measures in future studies. Multiple genes were associated with the brain volumes investigated in this, the largest and most geographically diverse genetic study to date. Genes identified were expressed in specific neural cell types that influence intracranial and subcortical brain volumes and are involved in autophagy, intracellular signalling and transport, organelle biogenesis and maintenance, or the aetiology of neurodegenerative disorders. Our findings point towards the generalisability of intracranial and subcortical brain volumes' polygenic scores to non-European ancestry individuals, suggesting a shared genetic basis of these brain volumes across diverse ancestral groups. We observed genetic overlap and putative causal genetic effects of intracranial and

subcortical brain volumes with neuropsychiatric conditions, including Parkinson's disease and ADHD. Overall, our findings advance the understanding of the brain's complex and polygenic genetic architecture, implicating multiple molecular pathways in human brain structure and suggesting that multiple genetic variants of small effect size are likely to be involved in the development of specific brain volumes. These studies also facilitate our understanding of shared genetic pathways underlying the aetiology of brain disorders and the formation and adaptation of the human brain.

Acknowledgments

We thank all of the study participants for contributing to this research. Full acknowledgements and grant support details are provided in the Supplementary Note.

Author contributions

Core analysis and writing group: LMGM, AIC, SDT, JAR, ZC, BLM, KLG, JGT, PMT, CLS, SEM, MER.

Cohort PIs: IA, DA, OAA, AA, DAB, MPMB, DIB, HB, JKB, WC, VC, SC, BC, AMD, GIdZ, CDeC, CDep, SDes, SEh, TE, SEF, MF, BF, HJG, OG, VG, AKH, UKH, AH, SH, PJH, MI, MAI, EGJ, RSK, LJL, SML, HLe, PMe, AM, THM, MMN, NGM, PAN, JO, TP, ZP, BWJHP, BMP, PSS, PGS, AJS, RS, GS, SS, JWS, DJS, JNT, DvE, HvB, NJAv, DJV, MWV, AV, HW, DRW, MWW, TW, AW.

Imaging data collection: IA, SA, KA, MEB, ASB, VC, BC, FC, EJCdG, CDeC, SEr, TE, GF, IF, DAF, ALG, OGr, OG, VG, AKH, UKH, SH, BH, AJH, NH, MAI, CJJr, EGJ, RSK, SML, HLe, DCML, JM, VSM, KLM, PMe, THM, TWM, SM, PAN, JO, MP, BWJHP, GP, NR, PGS, GS, SS, SMS, HSS, DTT, JNT, MCV, DvE, NJAv, MWV, HW, JMW, MWW, WW, LTW, EW, TW, MPZ.

Contributed to the editing of the manuscript: IA, DA, JCB, HB, JKB, JGT, VC, CRKC, GD, EJCdG, PL.DJ, SDes, SEh, TE, GF, IF, SEF, AJF, CF, BF, HJG, KLG, NAG, OG, UKH, DPH, SH, JJH, NH, MAI, JCI, NJ, MJK, SML, PHL, HLe, HL, WTL, ML, AFM, KAM, VSM, AM, BLM, THM, WJN, MMN, PAN, TP, BWJHP, BMP, JIR, PSS, CLS, SIT, AJS, GS, ES, SS, LSh, SMS, HSS, DJS, JLS, SIT, PMT, AT, JNT, MWV, JMW, DRW, MWW, LTW, TW, NMS, NGM, JR, LMGM, AIC, MER.

Genetic data collection: SA, PA, LA, RB, VC, SC, EJCdG, PL.DJ, CDep, SDes, SD, SEr, TE, SEF, AJF, CF, RCG, OG, VG, AKH, BH, GH, MAI, EGJ, SLH, DCML, JM, KAM, PMe, THM, TWM, MMN, PAN, BWJHP, BMP, BP, MDR, JIR, AJS, PRS, MSc, SS, LSh, SMS, HSS, VMS, NJAv, JV, HW.

Imaging data analysis: SA, KA, RMB, OTC, MC, QC, CRKC, BC, FC, CDeC, SDes, SEh, SEr, GF, IF, TG, ALG, OGr, NAG, AKH, UKH, DPH, SH, DFH, AJH, NJ, RK, DCML, PMa, AFM, KLM, SM, KN, WJN, PAN, JO, SLR, RR, GVR, PGS, CLS, LSc, DTT, MCV, DvE, NJAv, LNV, HW, JMW, WW, LTW, AW, MPZ.

Genetic data analysis: LA, JCB, JGT, RMB, QC, CRKC, SC, EJCdG, PL.DJ, SDeb, SDes, SD, SEh, MF, TG, KLG, NAG, DPH, EH, MK, MJK, SLH, PHL, SL, DCML, ML, YM, BLM, BM, KN, SLR, GVR, PGS, MSa, CLS, RS, PRS, MSc, LSh, JS, AVS, DvE, DvdM, CW, JY, LMGM, AIC, MER.

Tables

Table 1. Summary of GWAS meta-analysis results per subcortical brain volume and intracranial volume (ICV).

Brain volume	Number of genome-wide significant loci	h2SNP (SE)	Intercept (SE)	Attenuation ratio (SE)
Brainstem	96	0.35 (0.03)	1.00 (0.01)	< 0
ICV	83	0.28 (0.02)	1.00 (0.02)	0.006 (0.04)
Caudate nucleus	78	0.27 (0.02)	1.00 (0.01)	0.001 (0.03)
Putamen	71	0.29 (0.03)	1.00 (0.01)	0.006 (0.03)
Hippocampus	47	0.21 (0.02)	1.00 (0.01)	0.001 (0.04)
Ventral diencephalon	36	0.33 (0.03)	1.00 (0.01)	0.01 (0.04)
Thalamus	35	0.22 (0.01)	1.00 (0.01)	< 0
Globus pallidus	32	0.22 (0.02)	1.00 (0.01)	< 0
Nucleus accumbens	29	0.21 (0.01)	1.00 (0.01)	0.003 (0.03)
Amygdala	22	0.17 (0.01)	1.00 (0.01)	< 0
Total	529	NA	NA	NA

Number of genome-wide loci is reported at the common genome-wide significance threshold (p-value < $5x10^{-8}$, r^2 threshold to define independent significant loci ≥ 0.6, second r^2 threshold to define lead loci ≥ 0.05). A full list of independent significant loci is reported in Supplementary Table 1.

Figures

Figure 1. Meta-analyses results overview. Phenogram illustrating loci associated with each of the brain volumes under study at the common genome-wide significance threshold (*p*-value < $5x10^{-8}$). ICV = Intracranial volume, A = Left Hemisphere interior, B = Left Hemisphere exterior, C = Right hemisphere interior, D = Right hemisphere exterior, E = Both hemispheres upper. The p-values referenced here correspond to a two-tailed Z-test test as implemented in MTAG.

Figure 2. Polygenic prediction in the ABCD cohort.

Barplots show the variance explained by intracranial and subcortical brain volume polygenic scores using the *SBayesR* approach with a linear mixed effects model implemented in GCTA for the whole sample (N = 10,440), and individuals of European (N = 5,267), Non-European (N = 5,173), African-only (N = 1,833), and Asian-only (N = 152) ancestry. The *p*-value of the association is shown at the top of each bar; those with an asterisk (*) were significant after Bonferroni multiple testing correction [0.05 / 50 [total number of tests] = $1x10^{-3}$]. Non-European ancestry individuals include, but are not limited to, African-only and Asian-only ancestries as individuals with admixed ancestry were also included. P-values in this figure correspond to wald-tests (2-sided) derived from the linear mixed model results.

Figure 3. Genetic overlap with neuropsychiatric traits and disorders

Heatmap depicting genetic correlations (rG) of intracranial and subcortical brain volumes with complex human phenotypes. **p*-value < 0.05; ***p*-value significant after Bonferroni multiple testing correction (0.05 / 320 [total number of genetic correlation tests] = $1.56x10^{-4}$). Genetic correlations were estimated using LD score regression. P-values correspond to chi-squared tests with one degree of freedom as implemented in LD score regression.

Figure 4. Genetic structure of subcortical brain volumes

Path diagram of a three-factor model estimated with genomic structural equation modelling. Blue rectangles represent the genetic component of each subcortical brain volume. Green circles represent latent factors. Standardised path coefficients are presented.

References

1. Thompson, P. M. *et al.* ENIGMA and global neuroscience: A decade of large-scale studies of the brain in health and disease across more than 40 countries. *Transl. Psychiatry* **10**, 1–28 (2020).
2. Satizabal, C. L. *et al.* Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat. Genet.* **51**, 1624–1636 (2019).
3. Xu, H. *et al.* Subcortical Brain Abnormalities and Clinical Relevance in Patients With Hemifacial Spasm. *Front. Neurol.* **10**, 1383 (2019).
4. van Schouwenburg, M. R., den Ouden, H. E. M. & Cools, R. The human basal ganglia modulate frontal-posterior connectivity during attention shifting. *J. Neurosci.* **30**, 9910–9918 (2010).
5. Bickart, K. C., Wright, C. I., Dautoff, R. J., Dickerson, B. C. & Barrett, L. F. Amygdala volume and social network size in humans. *Nat. Neurosci.* **14**, 163–164 (2011).
6. Palomero-Gallagher, N. & Amunts, K. A short review on emotion processing: a lateralized network of neuronal networks. *Brain Struct. Funct.* **227**, 673–684 (2021).
7. Krabbe, K. *et al.* Increased intracranial volume in Parkinson's disease. *J. Neurol. Sci.* **239**, 45–52 (2005).
8. Tate, D. F. *et al.* Intracranial volume and dementia: some evidence in support of the cerebral reserve hypothesis. *Brain Res.* **1385**, 151 (2011).
9. Klein, M. *et al.* Genetic Markers of ADHD-Related Variations in Intracranial Volume. *Am. J. Psychiatry* **176**, (2019).
10. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).

11. Sønderby, I. E. *et al.* Dose response of the 16p11.2 distal copy number variant on intracranial volume and basal ganglia. *Mol. Psychiatry* **25**, 584 (2020).
12. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229 (2015).
13. Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. *Science* **367**, (2020).
14. Nawaz, M. S. *et al.* Thirty novel sequence variants impacting human intracranial volume. *Brain Commun* fcac271 (2022).
15. Adams, H. H. *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat. Neurosci.* **19**, (2016).
16. Hibar, D. P. *et al.* Novel genetic loci associated with hippocampal volume. *Nat. Commun.* **8**, 13624 (2017).
17. Hofer, E. *et al.* Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nat. Commun.* **11**, 1–16 (2020).
18. van der Meer, D. & Kaufmann, T. Mapping the genetic architecture of cortical morphology through neuroimaging: progress and perspectives. *Transl. Psychiatry* **12**, 1–11 (2022).
19. Loughnan, R. J. *et al.* Generalization of cortical MOSTest genome-wide associations within and across samples. *Neuroimage* **263**, (2022).
20. Park, S. M., Park, H. R. & Lee, J. H. MAPK3 at the Autism-Linked Human 16p11.2 Locus Influences Precise Synaptic Target Selection at Drosophila Larval Neuromuscular Junctions. *Mol. Cells* **40**, 151 (2017).
21. Choi, C. Y., Vo, M. T., Nicholas, J. & Choi, Y. B. Autophagy-competent mitochondrial translation elongation factor TUFM inhibits caspase-8-mediated apoptosis. *Cell Death Differ.* **29**, (2022).
22. Lee, S. & Choi, I. Expression patterns and biological function of Specc1 during mouse

preimplantation development. *Gene Expr. Patterns* **41**, (2021).

23. Kaku, H. & Rothstein, T. L. FAIM Is a Non-redundant Defender of Cellular Viability in the Face of Heat and Oxidative Stress and Interferes With Accumulation of Stress-Induced Protein Aggregates. *Frontiers in Molecular Biosciences* **7**, (2020).
24. Redenšek, S., Trošt, M. & Dolžan, V. Genetic Determinants of Parkinson's Disease: Can They Help to Stratify the Patients Based on the Underlying Molecular Defect? *Front. Aging Neurosci.* **9**, 20 (2017).
25. Ham, S. *et al.* Hydrocortisone-induced parkin prevents dopaminergic cell death via CREB pathway in Parkinson's disease model. *Sci. Rep.* **7**, (2017).
26. Cheng, W.-W., Zhu, Q. & Zhang, H.-Y. Identifying Risk Genes and Interpreting Pathogenesis for Parkinson's Disease by a Multiomics Analysis. *Genes* **11**, (2020).
27. Inda, C. *et al.* cAMP-dependent cell differentiation triggered by activated CRHR1 in hippocampal neuronal cells. *Sci. Rep.* **7**, (2017).
28. Jerber, J. *et al.* Population-scale single-cell RNA-seq profiling across dopaminergic neuron differentiation. *Nat. Genet.* **53**, 304–312 (2021).
29. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nature Human Behaviour* **3**, 513–525 (2019).
30. Liu, N. *et al.* Cross-ancestry genome-wide association meta-analyses of hippocampal and subfield volumes. *Nat. Genet.* **55**, 1126–1137 (2023).
31. Roshchupkin, G. V. *et al.* Heritability of the shape of subcortical brain structures in the general population. *Nat. Commun.* **7**, 13738 (2016).
32. Rentería, M. E. *et al.* Genetic architecture of subcortical brain regions: common and

region-specific genetic contributions. *Genes Brain Behav.* **13**, 821–830 (2014).

33. Smith, S. M. *et al.* An expanded set of genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nat. Neurosci.* **24**, 737–745 (2021).
34. Angeles, F.-G. M., Palacios-Bote, R., Leo-Barahona, M. & Mora-Encinas, J. P. Anatomy of the brainstem: a gaze into the stem of life. *Semin. Ultrasound CT MR* **31**, (2010).
35. Lanciego, J. L., Luquin, N. & Obeso, J. A. Functional Neuroanatomy of the Basal Ganglia. *Cold Spring Harb. Perspect. Med.* **2**, (2012).
36. Javed, N. & Cascella, M. Neuroanatomy, Globus Pallidus. in *StatPearls [Internet]* (StatPearls Publishing, 2023).
37. Young, C. B., Reddy, V. & Sonne, J. Neuroanatomy, Basal Ganglia. in *StatPearls [Internet]* (StatPearls Publishing, 2022).
38. Yang, Y. & Wang, J.-Z. From Structure to Behavior in Basolateral Amygdala-Hippocampus Circuits. *Front. Neural Circuits* **11**, 86 (2017).
39. Walton, E. *et al.* Exploration of Shared Genetic Architecture Between Subcortical Brain Volumes and Anorexia Nervosa. *Mol. Neurobiol.* **56**, (2019).
40. Schmaal, L. *et al.* Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol. Psychiatry* **21**, 806–812 (2015).
41. García-Marín, L. M. *et al.* Large-scale genetic investigation reveals genetic liability to multiple complex traits influencing a higher risk of ADHD. *Sci. Rep.* **11**, 1–9 (2021).
42. Bivol, S. *et al.* Australian Parkinson's Genetics Study (APGS): pilot (n=1532). *BMJ Open* **12**, e052032 (2022).
43. Hoogman, M. *et al.* Subcortical brain volume differences of participants with ADHD across the lifespan: an ENIGMA collaboration. *The lancet. Psychiatry* **4**, 310 (2017).

44. Müller, M. *et al.* YIP1 family member 4 (YIPF4) is a novel cellular binding partner of the papillomavirus E5 proteins. *Sci. Rep.* **5**, 1–14 (2015).

45. Zhang, C. *et al.* Genomic identification and expression analysis of nuclear pore proteins in Malus domestica. *Sci. Rep.* **10**, (2020).

46. Mao, Q. *et al.* KTN1 Variants Underlying Putamen Gray Matter Volumes and Parkinson's Disease. *Front. Neurosci.* **14**, (2020).

47. Yao, S. *et al.* A transcriptome-wide association study identifies susceptibility genes for Parkinson's disease. *npj Parkinson's Disease* **7**, 1–8 (2021).

48. Xu, J. *et al.* Environmental profiles of urban living relate to regional brain volumes and symptom groups of mental illness through distinct genetic pathways. *medRxiv* 2022.09.08.22279549 (2022) doi:10.1101/2022.09.08.22279549.

49. Cursano, S. *et al.* A CRHR1 antagonist prevents synaptic loss and memory deficits in a trauma-induced delirium-like syndrome. *Mol. Psychiatry* **26**, 3778 (2021).

50. Kim, M. S. *et al.* Testing the generalizability of ancestry-specific polygenic risk scores to predict prostate cancer in sub-Saharan Africa. *Genome Biol.* **23**, 1–16 (2022).

51. Lamballais, S., Jansen, P. R., Labrecque, J. A., Ikram, M. A. & White, T. Genetic scores for adult subcortical volumes associate with subcortical volumes during infancy and childhood. *Hum. Brain Mapp.* **42**, (2021).

52. Wang, Y. *et al.* Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nat. Commun.* **11**, 3865 (2020).

53. Campos, A. I. *et al.* Boosting the power of genome-wide association studies within and across ancestries by using polygenic scores. *Nat. Genet.* **55**, (2023).

54. Aschard, H., Vilhjálmsson, B. J., Joshi, A. D., Price, A. L. & Kraft, P. Adjusting for heritable

covariates can bias effect estimates in genome-wide association studies. *Am. J. Hum. Genet.* **96**, (2015).

55. Voevodskaya, O. *et al.* The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease. *Front. Aging Neurosci.* **6**, (2014).

56. Unverdi, M. & Alsayouri, K. Neuroanatomy, Cerebellar Dysfunction. in *StatPearls [Internet]* (StatPearls Publishing, 2023).

Methods

Statistics

This study performed several statistical approaches including linear regression, linear mixed effects associations, genome-wide association studies, LD-score regression, bivariate gaussian mixture models, genomic structural equation modelling, and MTAG-based meta analysis of GWAS summary statistics. Each approach is described in detail below.

Cohorts and GWAS

ENIGMA and CHARGE. GWAS summary statistics for the MRI-derived volume of seven subcortical brain structures of interest (nucleus accumbens, amygdala, brainstem, caudate nucleus, globus pallidus, putamen, and thalamus) were obtained from the ENIGMA website following the application and approval of this project. These GWAS summary statistics are detailed elsewhere[2]. This compilation of GWAS summary statistics is the product of a meta-analysis including 48 European ancestry samples from the ENIGMA consortium[57], the CHARGE consortium[58] and the first release (N~8,312) of the UK Biobank neuroimaging traits. Individual cohorts conducted quality control on their genotypic data (including SNP and sample level quality for MAF, missingness and heterozygosity), and phenotypic data (including outlier screening and distribution checks) prior to imputation. GWAS followed standardized ENIGMA/CHARGE analysis plans. Quality control before the meta-analysis of these samples included removing SNPs with poor imputation quality, removal of non-common SNPs (minor allele frequency > 0.01) and SNPs with a low effective minor allele count (< 20) or not represented across the meta-analysis (i.e. present in less than 70% of the total sample size for the discovery GWAS). Furthermore, a sample size (Z-score) weighted meta-analysis was used, as cohorts used different methods for acquisition, processing and adjustment of GWAS. The UK Biobank sample was adjusted for total brain volume, whereas ENIGMA and CHARGE consortium data were adjusted for total ICV[2]. Results from a previous GWAS meta-analysis for ICV[15] and hippocampal volume[16] were obtained via a public access repository through application and approval (https://enigma.ini.usc.edu/research/download-enigma-gwas-results/). Strict MRI-scan protocol procedures were followed to ensure high data quality as described thoroughly elsewhere[2].

UK Biobank. We performed GWAS for intracranial and nine subcortical brain volumes with data from the UK Biobank[59]. The UK Biobank genotyping and phenotyping have been described elsewhere[60]. Briefly, our GWAS includes 36,095 participants of European ancestry passing standard quality control procedures as described elsewhere[60]. The subcortical brain structures included the nucleus accumbens, amygdala, brainstem, caudate nucleus, hippocampus, globus pallidus, putamen, thalamus, and ventral diencephalon. We also performed a GWAS on ICV. We excluded outlier measures that were at least four standard deviations from the mean. GWASs were performed using BOLT-LMM (v2.3.2)[61], which accounts for relatedness via a linear mixed model. This method includes a random effect with a variance-covariance structure specified by a genetic-relatedness matrix (GRM) derived from a subset of SNPs across the genome[61]. The GWAS was adjusted for genotyping array, sex, age, sex*age, age-squared, sex*age-squared, and the first 20 genetic principal components to adjust further for population stratification. We included the neuroimaging data collection site (Data Field 54) as a covariate in the model to account for potential bias due to the use of different scanners across data collection sites. GWASs for subcortical brain volumes were further adjusted for ICV. We excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60) from the analysis. Strict MRI-scan protocol procedures were followed to ensure high data quality as described thoroughly elsewhere[62].

In the present study, we did not have an independent sample to perform replication analyses. Nonetheless, we leveraged the total sample from the UK Biobank included in our meta-analyses to create two subsamples of N ~ 18,047. These subsamples were created by randomly splitting the main sample N = 36,095 into two sets of data. We used these subsamples to conduct GWAS for intracranial and subcortical brain volumes as an alternative replication method to compare GWAS findings between these samples and with the meta-analyses. We included the same covariates as described above and performed the same quality control procedure.

Throughout the main set of GWAS analyses and for the meta-analyses we included ICV as a covariate in the GWAS to account for inter-individual variation in subcortical brain volume due to head size differences, which is crucial when using samples including participants from different

age groups. However, as previous studies have suggested[54], adjusting for heritable covariates, such as ICV, could bias effect estimates in GWAS. Therefore, as a sensitivity analysis, we also performed GWAS in the full sample of the UK Biobank cohort (N = 36,095) as described above for nine subcortical brain volumes, but without including ICV as a covariate in the model. This allowed us to understand potential differences in GWAS with and without correcting for ICV.

ABCD. The ABCD study is a longitudinal resource that includes children aged nine and ten at recruitment[63]. Conducted in the United States, neuroimaging measures were obtained by the ABCD Data Analysis and Information (DAIC) and the Image Acquisition workgroups. Neuroimaging was performed across 21 sites using three different scanner types. Further information on image acquisition and postprocessing is available elsewhere[64,65]. Brain volumes analysed in this cohort included ICV, hippocampus, ventral diencephalon, brainstem, nucleus accumbens, caudate nucleus, thalamus, globus pallidus, amygdala, and putamen —volumes of the left and right measures (where relevant)— were averaged for each individual. We excluded outlier measures that were at least four standard deviations from the mean. Saliva samples were obtained at a baseline visit, and genotyping was performed using a Smokescreen array following standard DNA extraction protocols. Quality control removed genetic variants with a low call rate (less than 99% of the sample) and samples with a missing rate greater than 20% or conflicting identifiers. This quality-controlled dataset was imputed to the 1000G Phase 3 reference panel using the Michigan Imputation Server [66]. Imputed genotype probabilities were extracted from the imputed data using QCTOOL v2 (https://www.well.ox.ac.uk/~gav/qctool_v2/). PLINK v2 was used to generate a subset of genetic files as a genetic relatedness matrix (GRM) for GWAS analysis. Briefly, a random list of 500,000 variants passing QC (minor allele frequency >= 0.01; CR >= 0.9 and INFO >= 0.6) was generated and used to create a new set of *PLINK* files (.bed, .bim, .fam) from the imputed genotype probability files. Ancestry was inferred by projecting the ABCD samples onto the principal components of the 1000 Genomes project using PLINK v1.90b6.8 and the flag --pca-clusters (**Supplementary Figure 128**). The Euclidean distance between the centroids for the first three principal components of each 1000G super population and each sample was calculated using Python (v3.5). To assess the validity of this approach, a receiver operating characteristic (ROC) curve was used to investigate whether this distance (multiplied by -1) was able to classify samples according to self-reported white race (which could be considered

a proxy for European ancestry). Participants were deemed outliers (i.e., non-Europeans) if they were more than three standard deviations from the super population centroid. Importantly, this cutoff value was close to Youden's *J*, which could be considered (post-hoc) the optimal cutoff for binary classification (**Supplementary Figure 129**). The final GWAS included 5,267 participants of European ancestry who passed genetic and neuroimaging quality control. The GWAS was performed using BOLT-LMM (v2.3.2) adjusting for age, sex*age, age-squared, sex*age-squared, and the first 20 genetic principal components to adjust further for population stratification. We included the imaging device serial number under the variable name 'mri_info_deviceserialnumber' as a covariate in the model, as suggested in previous studies[64], to account for potential bias due to the use of different scanners across data collection sites. Subcortical volumes GWAS were further adjusted for ICV. We excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60) from the analysis. Strict MRI-scan protocol procedures were followed to ensure high data quality as described thoroughly elsewhere[67].

Intracranial and subcortical brain volumes GWAS meta-analyses

We performed a GWAS meta-analysis for each brain volume phenotype across the ENIGMA-CHARGE published summary statistics and the GWAS in the UK Biobank and ABCD performed here, yielding a total sample size of up to 74,898 unique participants of European ancestry across all samples (**Supplementary Table 32**). All participants included in the present study provided written informed consent and the investigators on the participating studies obtained approval from their institutional review board or equivalent organisation. Individual GWAS for subcortical brain volumes were adjusted for ICV, as this reduces inter-individual variation in subcortical brain volume simply due to head size differences[68]. The meta-analyses were performed using MTAG v1.0.8[69]. Meta-analyses were performed, assuming equal heritability and perfect genetic covariance. Independent loci for human intracranial and subcortical brain volumes were determined by combining lead SNPs for all brain volumes under study and performing a conservative clumping procedure in PLINK 1.9[70] ($p^1 = 1x10^{-8}$, $p^2 = 1x10^{-5}$, $r^2 = 1x10^{-3}$, kb = 1000). Independent genome-wide loci not reported in previous studies are claimed based on a comparison of the independent unique loci identified in the present study across intracranial and subcortical brain volumes with independent genome-wide significant loci for intracranial[14,15]

or subcortical brain[2,16,30] volumes reported in previous studies. We considered linkage-disequilibrium information in the definition of the independent genome-wide loci not reported in previous studies by performing a clumping procedure using PLINK 1.9[70] ($p^1 = 1x10^{-8}$, $p^2 = 1x10^{-5}$, $r^2 = 1x10^{-3}$, kb = 1000). We report results at the common genome-wide significance threshold. In addition, we performed multiple testing correction using matSpD to account for the total number of phenotypes as performed in previous studies[12]. We observed that the effective number of independent traits in our analysis was 8. Thus, we set a significance threshold of *p*-value < $5x10^{-8} / 8 = 6.25x10^{-9}$.

The imaging analysis and visualization of structural data in all cohorts was performed using the publicly available FreeSurfer[71] package tool (https://surfer.nmr.mgh.harvard.edu/) developed by the Laboratory for Computational Neuroimaging at the Athinoula A. Martinos Center for Biomedical Imaging. Details regarding border definition for specific brain structures is available on the wiki (https://surfer.nmr.mgh.harvard.edu/fswiki/FreeSurferWiki). In particular, the ventral diencephalon contains the following structures: the hypothalamus, basal forebrain, the sublenticular extended amygdala (SLEA), and a portion of the ventral tegmentum, which can also be considered a part of the midbrain[71,72]. These specific substructures do not overlap with the brainstem borders, which is constituted by the medulla oblongata, pons, midbrain and superior cerebellar peduncle[71].

The phenogram in **Figure 1** was created using the Ritchie Lab Visualization online tool (https://visualization.ritchielab.org/phenograms/plot). Subcortical brain images in **Figure 1** were created using publicly available tutorials (https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation and https://bookdown.org/u0243256/tbicc/freesurfer.html) in MATLAB (R2023b). The ENIGMA consortia also provides tutorials on the creation of brain-related figures (https://enigma-toolbox.readthedocs.io/en/latest/pages/12.visualization/index.html#subcortical-surface-visualization).

Functional annotation and gene prioritisation

We performed functional annotation and gene prioritisation analyses using MAGMA, eQTL mapping with TWAS, and by integrating single cell sequencing data with GWAS summary statistics.

First, we performed gene-based tests using MAGMA[73] v(1.08) as implemented in FUMA (v1.5.2)[74] (https://fuma.ctglab.nl/). The MAGMA method provides aggregate association *p*-values based on all variants within a gene and its regulatory region[73]. We applied a Bonferroni multiple-testing correction based on the total number of genes and accounted for the effective number of independent traits in our analysis, (0.05 / 17,708 [average number of tests per brain volume] * 8 [estimated number of independent phenotypes] = 2.26×10^{-5}).

Second, we conducted an in-depth analysis of genetically regulated gene expression using FUSION (http://gusevlab.org/projects/fusion/), a software tool for TWAS[75]. FUSION leverages SNP-gene expression associations to construct predictive linear models tailored to each gene. The model demonstrating superior predictive performance in cross-validation trials was subsequently employed for predictive applications within the GWAS. Available tissue specimens sourced from five distinct subcortical regions from GTEx v8 (specifically, Accumbens, Amygdala, Hypothalamus, Hippocampus, and Putamen) were included in the analysis. For this, we used a Mendelian randomization framework employing summary-data-based Mendelian Randomisation (SMR) v1.3.1[76] to assess gene expression in multiple cell lines across the nine subcortical and intracranial volumes (ICV). We also incorporated data from RNA splicing sequencing, based on single-tissue gene expression derived from the brain. We applied Bonferroni multiple testing correction and accounted for the effective number of independent traits in our analysis (0.05 / 1,308 [average number of annotations per brain volume] * 8 [estimated number of independent phenotypes] = 3.06×10^{-4}). Moreover, we utilised an eQTL dataset derived from 120 human fetal brains[77], employing the SMR method to identify genes involved in the development of subcortical brain structures. We applied Bonferroni multiple testing correction and accounted for the effective number of independent traits in our analysis (0.05 / 317 [average number of annotations per brain volume] * 8 [estimated number of independent phenotypes] = 1.26×10^{-3}).

Genes prioritised through MAGMA and FUSION analyses from single and multiple brain tissues were further assessed by integrating GWAS summary data with single-cell RNA-sequencing data, which included over 1 million cells at three different stages of the differentiation process. Single-cell RNA-Seq analysis was based on eQTL data of Jerber et al. from 215 human induced pluripotent stem cell lines as they progressed towards a midbrain neural-like fate[28]. This process encompasses the generation of dopaminergic neurons, serotonin transporters, astrocyte-like cells, ependymal cells, and neuron-differentiated clusters. We filtered results for those involving genes associated with intracranial or subcortical brain volumes across MAGMA and TWAS analyses. Then, we applied Bonferroni multiple testing correction technique, considering the effective number of independent traits in our analysis (0.05 / 337 [total number of gene-brain volume associations] * 8 [estimated number of independent phenotypes] = 1.19×10^{-3}).

As a sensitivity analysis, we sought to understand potential differences in GWAS for subcortical brain volumes with and without correcting for ICV. Therefore, we performed gene-based tests using MAGMA[73] v(1.08) as implemented in FUMA[74] for GWAS in the UK Biobank cohort with and without adjusting for ICV. For each set of GWAS summary statistic (i.e., with and without adjusting for ICV), we applied Bonferroni multiple testing correction technique, considering the effective number of independent traits in our analysis (0.05 / 1097 [total number of gene-brain volume associations] * 8 [estimated number of independent phenotypes] = 3.64×10^{-4}).

SNP-based heritability and genetic correlations

LD score regression (LDSC)[78] was used to estimate the heritability for each subcortical brain structure. Briefly, this method leverages the expected relationship between the linkage-disequilibrium variant tags and their expected degree of association with a given phenotype to estimate the heritability. It distinguishes between confounding bias and polygenicity[78]. We processed our meta-analysis results using the *munge* function from LDSC v.1.0.1 and performed LD score regression to estimate the percentage of variance explained by the SNPs in the meta-analysis.

The genetic correlation between a pair of phenotypes depicts the relationship of genetic effect sizes at mutual genetic variants across phenotypes[79]. In the present study, we used LD score

regression to perform genetic correlation analyses among subcortical brain structures and between complex human phenotypes, including neuropsychiatric disorders and anthropometric measurements, with subcortical brain structures. Details for the GWAS summary statistics for neuropsychiatric and subcortical brain structures are provided in the **Supplementary Table 33** and **Supplementary Methods**. These complex human phenotypes were selected based on criteria applied in previous studies by the ENIGMA consortium, which relies on the public availability of well-powered summary statistics of previously reported brain-related phenotypes and anthropometric measurements[2,15]. These criteria are limited and restricted in the present study by the data transfer agreement with the CHARGE cohort, for which we are not allowed to leverage CHARGE data to investigate any relationships involving substance-related disorders and cognitive or intelligence-related phenotypes. We accounted for multiple testing using Bonferroni correction (0.05 / 320 [total number of genetic correlation tests] = 1.56×10^{-4}).

As a sensitivity analysis, we sought to understand potential differences in GWAS for subcortical brain volumes with and without correcting for ICV. Therefore, we estimated the genetic correlation between the GWAS for subcortical brain volumes in the UK Biobank cohort with and without adjusting for ICV. In addition, we estimated genetic correlations for both sets of GWAS summary statistics (i.e., with and without adjusting for ICV) with complex human phenotypes.

Pairwise GWAS

We leveraged the pairwise-GWAS (v.0.3.6) method[80] to identify segments of the genome with genomic variants influencing the aetiology of a brain volume and a human complex phenotype. For each pair of genetically correlated phenotypes after multiple testing correction according to our LD score regression results, we conducted GWAS-PW analyses. This method splits the genome into 1703 independent segments and, for each segment, GWAS-PW estimates the posterior probability of association (PPA) for four different models. These models include (i) the genomic segment is uniquely associated to phenotype A, (ii) the region is uniquely associated to phenotype B, (iii) the segment of the genome is influencing the aetiology of both phenotypes through the same genetic variants, and (iv) the genomic segment is involved in the aetiology of both phenotypes via different genetic variants. We provide findings for segments of the genome where model three (the genomic segment is influencing the aetiology of both phenotypes through

the same genetic variants) had a PPA > 0.5, given that this threshold has been used in previous studies[81,82].

Bivariate MiXeR

We conducted bivariate MiXeR analyses using MiXeR v1.3[83] to quantify polygenicity among the nine subcortical brain volumes under study. This analysis has been thoroughly described elsewhere[83]. Briefly, MiXeR leverages GWAS summary statistics and a univariate gaussian mixture model to estimate the degree of polygenicity (irrespective of genetic correlation), which is commonly referred to as the number of trait-influencing genetic variants. Then, with a bivariate gaussian mixture model, the additive genetic effect of four components is estimated for every pair of phenotypes: (i) genetic variants that do not influence either phenotype, (ii) genetic variants that only influence phenotype A (iii) genetic variants that only influence phenotype B, and (iv) genetic variants that influence both phenotypes[83]. Thus, MiXeR provides information about the genetic associations between two complex phenotypes as it estimates the total number of shared and phenotype-specific causal variants.

Genetic factor analyses

To examine genetic clustering of the nine subcortical brain structures we conducted exploratory factor analyses (EFA) based upon the LDSC-derived genetic correlation matrix. The R (v3.5.1) package 'psych' was used to conduct the EFAs, with a maximum likelihood extraction method and oblimin rotation method. The factor models identified in the EFA (retaining factor loadings > 0.25) were subsequently carried forward in a confirmatory factor analysis (CFA) in genomic SEM. This was done to assess the fit of the factor model to the data while taking into account uncertainty in covariance estimates. The default diagonally weighted least squares estimator was used.

Potential causal genetic effects

We leveraged the latent causal variable (LCV)[84] and Latent Heritable Confounder Mendelian Randomisation (LHC-MR v0.0.0.9000)[85] methods to investigate potential causal genetic effects between brain volumes under study and those complex human traits that displayed a statistically significant genetic correlation after Bonferroni multiple testing correction.

We employed the LCV method, which has been thoroughly described elsewhere[79,84], to assess whether the genetic correlations identified in the present study could be explained by putative causal genetic effects and accounted for multiple testing using a Bonferroni correction [0.05 / 16 [total number of genetic causal proportion tests in the present study] = $3.13x10^{-3}$]. Advantages of the LCV method include that (i) is it less susceptible to confounding by horizontal pleiotropic effects, (ii) it leverages aggregated information across the entire genome (i.e., full genome-wide data) to increase statistical power, and (iii) it is robust to sample overlap[84].

In the LCV method, the sign of the GCP parameter denotes the direction of potential causal genetic effects[84]. The GCP parameter ranges from -1 to 1, where GCP = 1 suggests full putative causal genetic effects of phenotype A on phenotype B. Conversely, GCP = -1 suggests full putative causal genetic effects of phenotype B on phenotype A. Moreover, a GCP = 0 implies the detection of horizontal pleiotropy, suggesting that an intervention on one phenotype would not affect the other due to the absence of causal genetic effects. Overall, to interpret LCV findings one must consider three important factors: (i) the magnitude of the genetic correlation, (ii) the GCP estimate, and (iii) the direction (positive or negative) of the GCP estimate[84,86–88].

LHC-MR leverages full GWAS summary statistics (not only genome-wide independent loci like traditional MR methods) to investigate potential causal genetic effects between a pair of genetically correlated phenotypes. LHC-MR has been reported to improve statistical power to estimate bi-directional putative causal genetic effects, direct heritabilities, and confounder effects while accounting for sample overlap. LHC-MR has been suggested to outperform a number of traditional MR methods[85]. Full details for the LHC-MR method are described elsewhere[85]. We accounted for multiple testing using a Bonferroni correction [0.05 / 32 [total number of LHC-MR tests in the present study] = $1.56x10^{-3}$]. We performed LHC-MR analyses with R (v3.5.1).

Polygenic scores estimation and association analyses

We performed the meta-analysis again but without the ABCD cohort to ensure sample independence and test polygenic prediction in European (N = 5,267), non-European (N = 5,173), African-only (N = 1,833), Asian-only (N = 152), and all samples (N = 10,440). Non-European

ancestry individuals include, but are not limited to african-only and asian-only ancestries as individuals with admixed ancestry were also considered. To avoid bias due to the correlation between SNPs arising from linkage-disequilibrium (LD), a Bayesian analysis was used to approximate the results of a conditional GWAS (i.e. one estimating the effect for all SNPs simultaneously). This was performed using *SBayesR*[89] implemented within the *Genome-wide Complex Trait Bayesian analysis* (GCTB v2.0) software tool[90]. Polygenic scores for intracranial and subcortical brain volumes were estimated by multiplying the multivariate effect size (obtained from *SBayesR*) times the allelic dosage of the effect allele and summing across all loci for each participant. Only SNPs passing quality control (minor allele frequency > 0.01, call rate > 0.9 and imputation score > 0.6) were included in the derived polygenic scores. To test for the association between intracranial and subcortical brain volumes polygenic scores with their corresponding phenotype and estimate the percentage of phenotypic variance explained, we performed a linear mixed effects model, in GCTA version 1.91.7 beta1, with a random effect and with a variance-covariance specified by a genetic relatedness matrix to account for cryptic relatedness among participants of the ABCD cohort. The results were plotted in Python (v3.5) using *seaborn*, *matplotlib* and in-house scripts. Sensitivity analyses assessed whether differential variance explained within the ABCD cohort was due to differential ancestry, sample size differences, or cryptic relatedness. These analyses consisted of (i) using a clumping and thresholding approach to derive polygenic scores with a linear mixed effects model implemented in GCTA to perform the prediction; (ii) performing the association analyses using a multivariate linear regression in Python *(*v3.5), and the library *statsmodels*. Additional covariates included in the model were sex, age, and the first twenty genetic ancestry components to adjust for population stratification. The percentage of variance explained was estimated as the difference in R^2 between the full model (i.e., including the polygenic scores) and a reduced model including only covariates; and (iii) Using *SBayesR* derived polygenic scores to perform associations analyses with multivariate linear regressions among participants of European ancestry including ICV as one of the covariates.

Ethics statement

Our study is based on meta-analysis of previously published, publicly available data for which appropriate site-specific Institutional Review Boards and ethical review at local institutions have

previously approved the use of these data. For full-details on the institutions that have approved the use of these data please refer to the 'Acknowledgments' section.

Data availability

Detailed information on how to access publicly available GWAS summary data from the ENIGMA and CHARGE consortia is reported on their corresponding publications[2,12,15]. Researchers can access individual-level data from the UKB and ABCD cohorts following the corresponding data application procedures. Work performed using UKB data was done under application 25331. Full genome-wide summary statistics generated in the present study are available at the ENIGMA website (http://enigma.ini.usc.edu/research/download-enigma-gwas-results).

Code availability

No custom code was used in this study. Publicly available software tools were used to perform genetic analyses and are referenced throughout the manuscript.

Competing interests

IA received speaker's honorarium Lundbeck; OAA is a consultant to Cortechs.ai and Precision Health, speaker's honorarium from Lundbeck, Janssen, Otsuka, Sunovion; HB is an Advisory Board Member or Consultant to Biogen, Eisai, Eli Lilly, Roche, Skin2Neuron, Cranbrook Care and Montefiore Homes; CRKC has received past partial research support from Biogen, Inc. (Boston, USA) for work unrelated to the topic of this manuscript; AMD is the Principal Investigator of a research agreement between General Electric Healthcare and the University of California, San Diego (UCSD); he is a Founder of and hold equity in CorTechs Labs, Inc. I am a member of the Scientific Advisory Board of Human Longevity, Inc., and the Mohn Medical Imaging and Visualization Center in Bergen, Norway. The terms of these arrangements have been reviewed and approved by UCSD in accordance with its conflict-of-interest policies; BF has received educational speaking fees from Medice; HJG has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm, Servier and Janssen Cilag as well as research funding from Fresenius Medical Care; DPH is a full time employee of Genentech, Inc; NH is a shareholder

various manufacturers of medical technology; AM-L has received consultant fees from Daimler und Benz Stiftung, EPFL Brain Mind Institute, Fondation FondaMental, Hector Stiftung II, Invisio, Janssen-Cilag GmbH, Lundbeck A/S, Lundbeckfonden, Lundbeck Int. Neuroscience Foundation, Neurotorium, MedinCell, The LOOP Zürich, University Medical Center Utrecht, University of Washington, Verein für Mentales Wohlbefinden, von Behring-Röntgen-Stiftung; speaker fees from Ärztekammer Nordrhein, Caritas, Clarivate, Dt. Gesellschaft für Neurowissenschaftliche Begutachtung, Gentner Verlag, Landesärztekammer Baden-Württemberg, LWL Bochum, Northwell Health, Ruhr University Bochum, Penn State University, Society of Biological Psychiatry, University Prague, Vitos Klinik Rheingau and editorial and/or author fees fromAmerican Association for the Advancement of Science, ECNP, Servier Int., Thieme Verlag; WN is founder of Quantib BV and was scientific lead of Quantib BV until Jan 31, 2023; MMN has received fees for membership in an advisory board from HMG Systems Engineering GmbH (Fürth, Germany), for membership in the Medical-Scientific Editorial Office of the Deutsches Ärzteblatt, and for serving as a consultant for EVERIS Belgique SPRL in a project of the European Commission (REFORM/SC2020/029). MMN receives salary payments from Life & Brain GmbH and holds shares in Life & Brain GmbH. All these concerned activities outside the submitted work; BMP serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson; AJS receives support from multiple NIH grants (P30 AG010133, P30 AG072976, R01 AG019771, R01 AG057739, U19 AG024904, R01 LM013463, R01 AG068193, T32 AG071444, U01 AG068057, U01 AG072177, and U19 AG074879). He has also received support from Avid Radiopharmaceuticals, a subsidiary of Eli Lilly (in kind contribution of PET tracer precursor); Bayer Oncology (Scientific Advisory Board); Eisai (Scientific Advisory Board); Siemens Medical Solutions USA, Inc. (Dementia Advisory Board); NIH NHLBI (MESA Observational Study Monitoring Board); Springer-Nature Publishing (Editorial Office Support as Editor-in-Chief, Brain Imaging and Behavior); MSreceived funding from Pfizer Inc. for a project not related to this research; ES received speaker fees from bfd buchholz fachinformationsdienst gmbh; PMT receives partial research support from Biogen, Inc., for research unrelated to this manuscript; MWW serves on Editorial Boards for Alzheimer's & Dementia, and the Journal for Prevention of Alzheimer's disease. He has served on Advisory Boards for Acumen Pharmaceutical, Alzheon, Inc., Cerecin, Merck Sharp & Dohme Corp., and NC Registry for Brain Health. He also serves on the USC ACTC grant which receives funding from Eisai for the AHEAD study. MWW has provided

consulting to Boxer Capital, LLC, Cerecin, Inc., Clario, Dementia Society of Japan, Eisai, Guidepoint, Health and Wellness Partners, Indiana University, LCN Consulting, Merck Sharp & Dohme Corp., NC Registry for Brain Health, Prova Education, T3D Therapeutics, University of Southern California (USC), and WebMD. MWW has acted as a speaker/lecturer for China Association for Alzheimer's Disease (CAAD) and Taipei Medical University, as well as a speaker/lecturer with academic travel funding provided by: AD/PD Congress, Cleveland Clinic, CTAD Congress, Foundation of Learning; Health Society (Japan), INSPIRE Project; U. Toulouse, Japan Society for Dementia Research, and Korean Dementia Society, Merck Sharp & Dohme Corp., National Center for Geriatrics and Gerontology (NCGG; Japan), University of Southern California (USC). MWW holds stock options with Alzeca, Alzheon, Inc., ALZPath, Inc., and Anven. MWW received support for his research from the following funding sources: National Institutes of Health (NIH)/NINDS/National Institute on Aging (NIA), Department of Defense (DOD), California Department of Public Health (CDPH), University of Michigan, Siemens, Biogen, Hillblom Foundation, Alzheimer's Association, Johnson & Johnson, Kevin and Connie Shanahan, GE, VUmc, Australian Catholic University (HBI-BHR), The Stroke Foundation, and the Veterans Administration; AIC is currently employed by the Regeneron Genetics Center, a wholly-owned subsidiary of Regeneron Pharmaceuticals, Inc., and may hold Regeneron stock or stock options. All other authors declare no competing interests.

Methods-only references

57. Thompson, P. M. *et al.* The ENIGMA Consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain Imaging Behav.* **8**, 153–182 (2014).

58. Psaty, B. M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2**, 73–80 (2009).

59. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

60. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

61. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

62. Alfaro-Almagro, F. *et al.* Image processing and Quality Control for the first 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**, 400 (2018).

63. Volkow, N. D. *et al.* The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev. Cogn. Neurosci.* **32**, 4–7 (2018).

64. Hagler, D. J., Jr *et al.* Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage* **202**, 116091 (2019).

65. Casey, B. J. *et al.* The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).

66. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

67. Saragosa-Harris, N. M. *et al.* A practical guide for researchers and reviewers using the ABCD

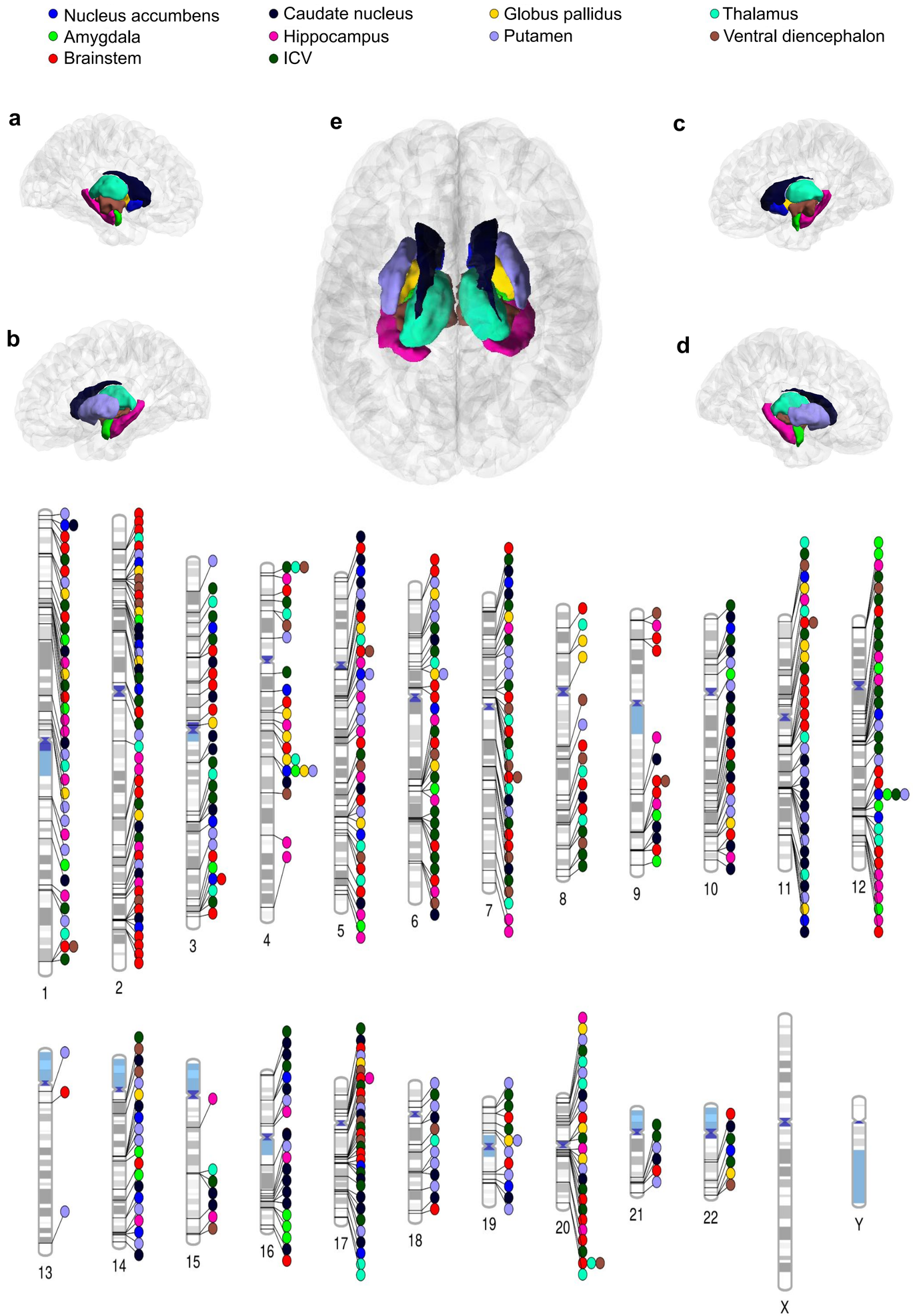
Study and other large longitudinal datasets. *Dev. Cogn. Neurosci.* **55**, (2022).

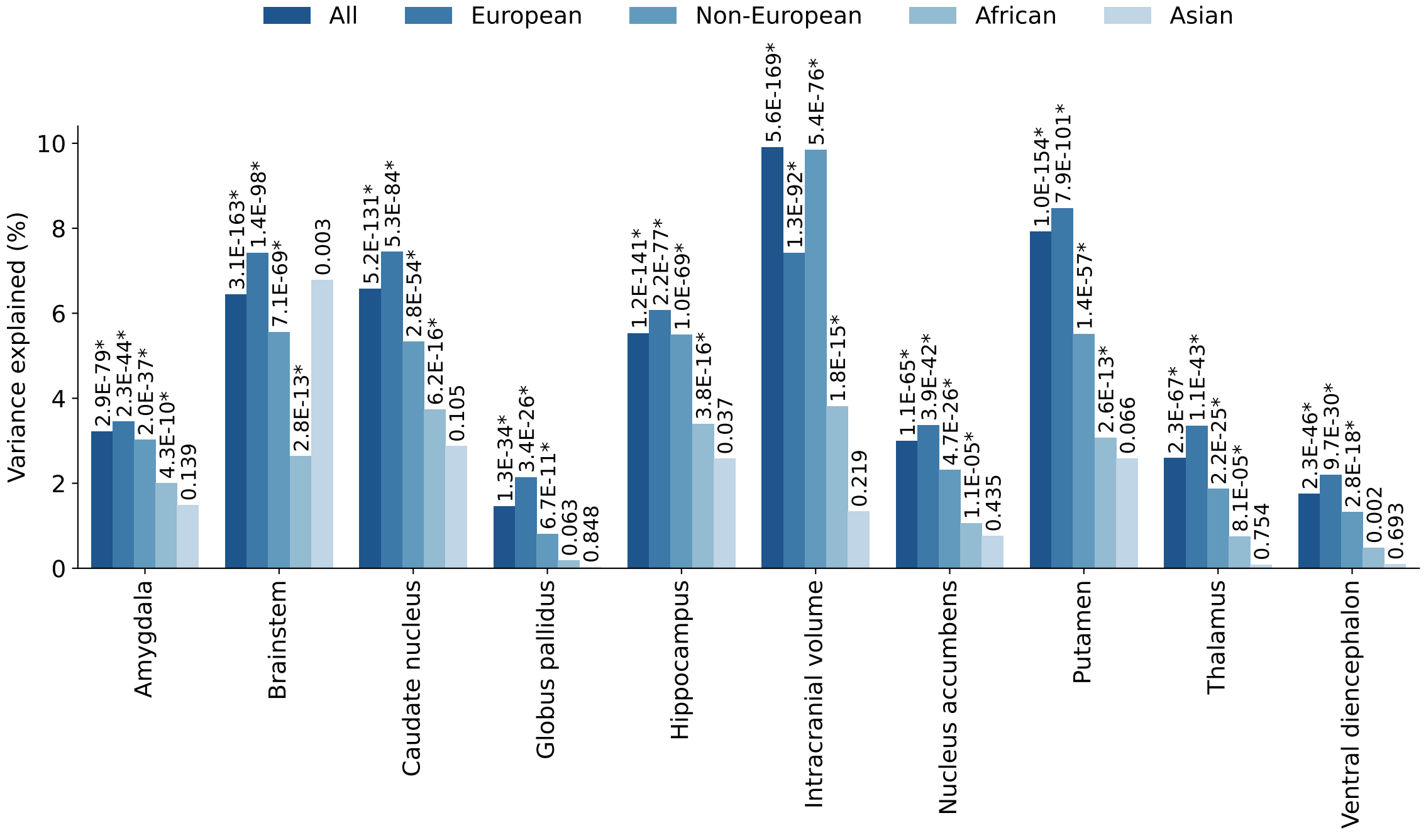
68. Crowley, S. *et al.* Considering Total Intracranial Volume and Other Nuisance Variables in Brain Voxel Based Morphometry in Idiopathic PD. *Brain Imaging Behav.* **12**, 1 (2018).
69. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
70. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559 (2007).
71. Fischl, B. FreeSurfer. *Neuroimage* **62**, (2012).
72. Makris, N. *et al.* Decreased volume of the brain reward system in alcoholism. *Biol. Psychiatry* **64**, (2008).
73. de Leeuw, C. A., Mooij, J. M. & Heskes, T. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, (2015).
74. Watanabe, K., Taskesen, E. & van Bochoven, A. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1–11 (2017).
75. Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).
76. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, (2016).
77. O'Brien, H. E. *et al.* Expression quantitative trait loci in the developing human brain and their enrichment in neuropsychiatric disorders. *Genome Biol.* **19**, 1–13 (2018).
78. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
79. García-Marín, L. M., Campos, A. I., Martin, N. G., Cuéllar-Partida, G. & Rentería, M. E. Inference

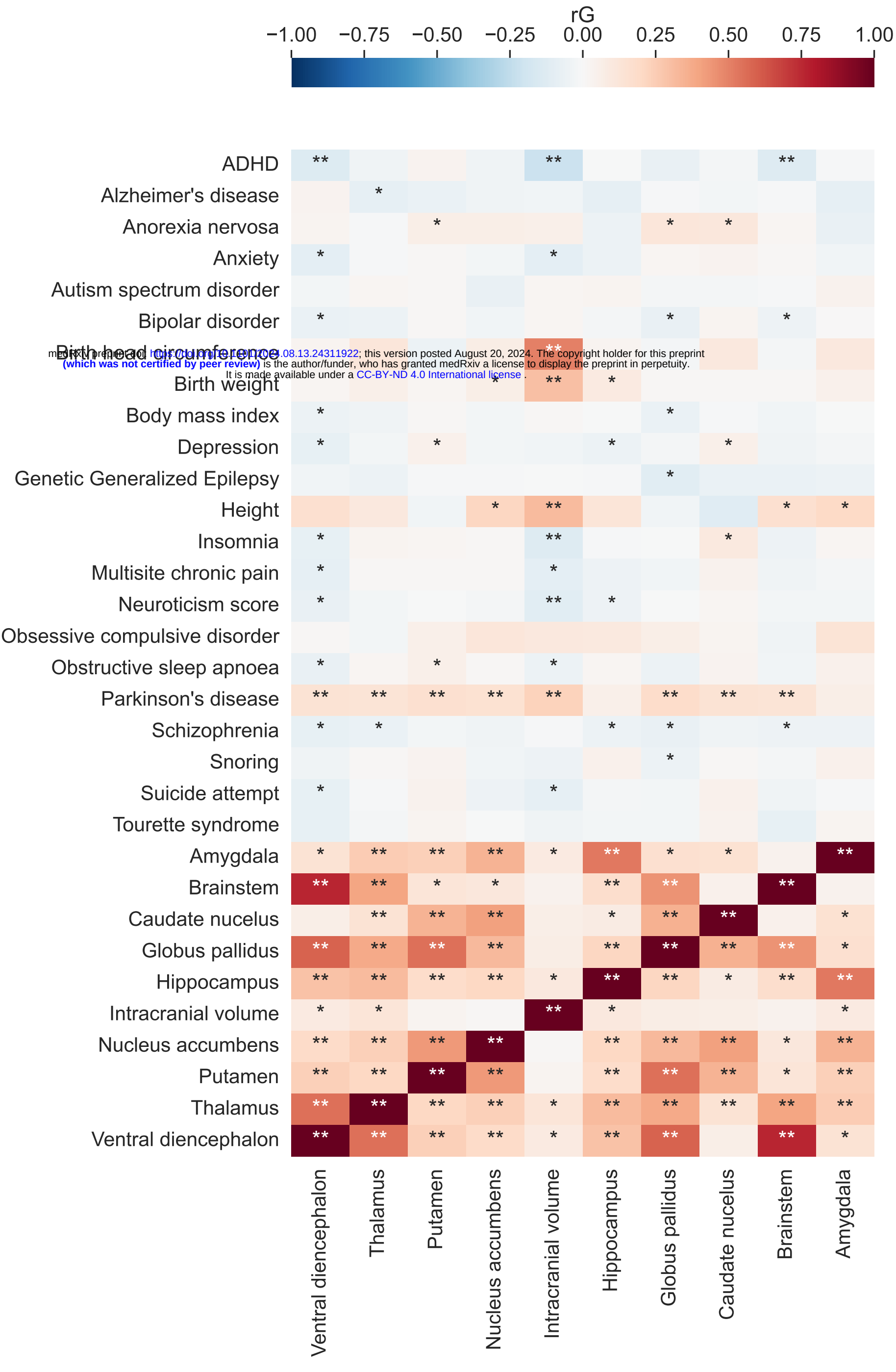
of causal relationships between sleep-related traits and 1,527 phenotypes using genetic data. *Sleep* **44**, zsaa154 (2020).

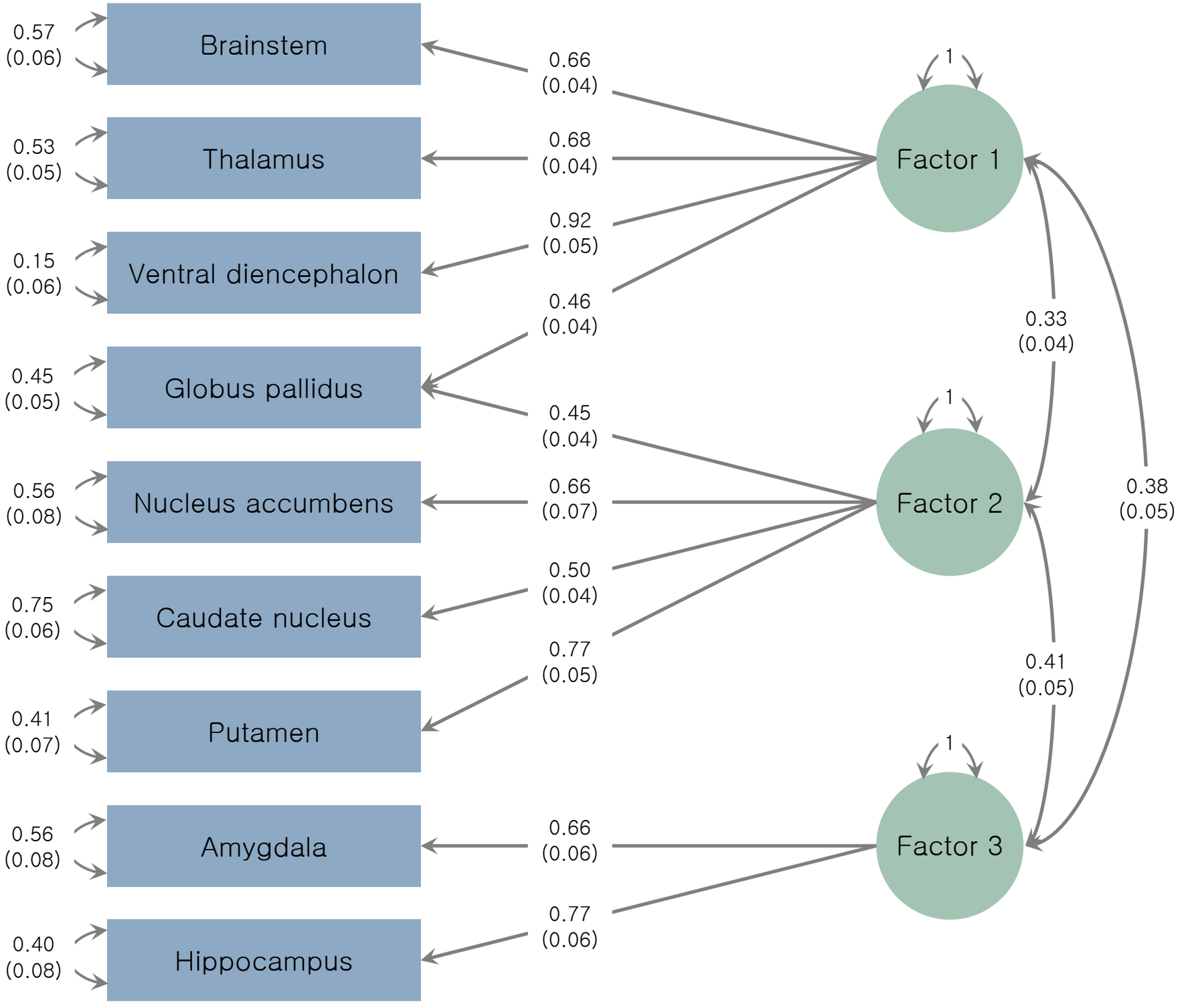
80. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**, (2016).
81. García-Marín, L. M. *et al.* Shared molecular genetic factors influence subcortical brain morphometry and Parkinson's disease risk. *npj Parkinson's Disease* **9**, 1–10 (2023).
82. Mitchell, B. L. *et al.* Elucidating the relationship between migraine risk and brain structure using genetic data. *Brain* **145**, 3214–3224 (2022).
83. Frei, O. *et al.* Bivariate causal mixture model quantifies polygenic overlap between complex traits beyond genetic correlation. *Nat. Commun.* **10**, 1–11 (2019).
84. O'Connor, L. J. & Price, A. L. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat. Genet.* **50**, 1728 (2018).
85. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *Nat. Commun.* **12**, 1–15 (2021).
86. Aman, A. M. *et al.* Phenome-wide screening of the putative causal determinants of depression using genetic data. *Hum. Mol. Genet.* **31**, 2887–2898 (2022).
87. García-Marín, L. M. *et al.* Phenome-wide screening of GWAS data reveals the complex causal architecture of obesity. *Hum. Genet.* **140**, (2021).
88. García-Marín, L. M., Campos, A. I., Martin, N. G., Cuéllar-Partida, G. & Rentería, M. E. Phenome-wide analysis highlights putative causal relationships between self-reported migraine and other complex traits. *J. Headache Pain* **22**, 1–8 (2021).
89. Lloyd-Jones, L. R. *et al.* Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat. Commun.* **10**, 5086 (2019).

90. Chung, W. Statistical models and computational tools for predicting complex traits and diseases. *Genomics Inform.* **19**, (2021).




All
European
Non-European
African
Asian
Variance explained (%)
0
2
4
6
8
10
Amygdala
2.9E-79*
2.3E-44*
2.0E-37*
4.3E-10*
0.139
Brainstem
3.1E-163*
1.4E-98*
7.1E-69*
2.8E-13*
0.003
Caudate nucleus
5.2E-131*
5.3E-84*
2.8E-54*
6.2E-16*
0.105
Globus pallidus
1.3E-34*
3.4E-26*
6.7E-11*
0.063
0.848
Hippocampus
1.2E-141*
2.2E-77*
1.0E-69*
3.8E-16*
0.037
Intracranial volume
5.6E-169*
1.3E-92*
5.4E-76*
1.8E-15*
0.219
Nucleus accumbens
1.1E-65*
3.9E-42*
4.7E-26*
1.1E-05*
0.435
Putamen
1.0E-154*
7.9E-101*
1.4E-57*
2.6E-13*
0.066
Thalamus
2.3E-67*
1.1E-43*
2.2E-25*
8.1E-05*
0.754
Ventral diencephalon
2.3E-46*
9.7E-30*
2.8E-18*
0.002
0.693




Brainstem
Thalamus
Ventral diencephalon
Globus pallidus
Nucleus accumbens
Caudate nucleus
Putamen
Amygdala
Hippocampus
0.57 (0.06)
0.53 (0.05)
0.15 (0.06)
0.45 (0.05)
0.56 (0.08)
0.75 (0.06)
0.41 (0.07)
0.56 (0.08)
0.40 (0.08)
0.66 (0.04)
0.68 (0.04)
0.92 (0.05)
0.46 (0.04)
0.45 (0.04)
0.66 (0.07)
0.50 (0.04)
0.77 (0.05)
0.66 (0.06)
0.77 (0.06)
Factor 1
Factor 2
Factor 3
1
1
1
0.33 (0.04)
0.38 (0.05)
0.41 (0.05)