Title page

Title: 1q21.1 distal copy number variants are associated with cerebral and cognitive alterations in humans

Authorships:

Ida E Sønderby, PhD 1-3, Dennis van der Meer, PhD 1,4, Clara Moreau, PhD 5,6, Tobias Kaufmann, PhD 1, G. Bragi Walters, B.Sc. 7,8, Maria Ellegaard, PhD 9, Abdel Abdellaoui, PhD 10,11, David Ames, MD 12,13, Katrin Amunts, PhD, MD 14,15, Micael Andersson, MSc 16,17, Nicola J Armstrong, PhD 18, Manon Bernard, BSc 19, Nicholas B Blackburn, PhD 20, John Blangero, PhD 20, Dorret I Boomsma, PhD 11,21,22, Henry Brodaty, PhD, MD 23,24, Rachel M. Brouwer, PhD 25, Robin Bülow, MD 26, Rune Bøen, MPhil 1-2, Wiepke Cahn, PhD, MD 25,27, Vince D Calhoun, PhD 28,29, Svenja Caspers, PhD, MD 14,30, Christopher RK Ching, PhD 31, Sven Cichon, PhD 14,32,33, Simone Ciufolini, PhD, MD 34, Benedicto Crespo-Facorro, PhD, MD 35,36, Joanne E Curran, PhD 20, Anders M Dale, PhD 37, Shareefa Dalvie, PhD 38, Paola Dazzan, Professor 39, Eco JC de Geus, PhD 10,21,22, Greig I de Zubicaray, PhD 40, Sonja MC de Zwarte, MSc 25, Sylvane Desrivieres, PhD 41, Joanne L Doherty, PhD 42,43, Gary Donohoe, PhD 44, Bogdan Draganski, MD 45, Stefan Ehrlich, PhD, MD 46, Else Eising, PhD 47, Thomas Espeseth, PhD 48, Kim Fejgin, PhD 49, Simon E Fisher, DPhil 47,50, Tormod Fladby, PhD, MD 51,52, Oleksandr Frei, PhD 1, Vincent Frouin, MSc 53, Masaki Fukunaga, PhD 54,55, Thomas Gareau, 53, Tian Ge, PhD 56,57,58, David C Glahn, PhD 59,60,61, Hans J Grabe, MD 62,63, Nynke A Groenewold, PhD 38, Ómar Gústafsson, PhD 7, Jan Haavik, PhD, MD 64,65, Asta K Haberg, PhD, MD 66,67, Jeremy Hall, MA, MB, Bchir, MPhil, PhD, MRCPsych 42,68, Ryota Hashimoto, PhD, MD 69,70, Jayne Y Hehir-Kwa, PhD 71, Derrek P Hibar, PhD 72, Manon H.J. Hillegers, PhD, MD 73, Per Hoffmann, PhD 74,33, Laurena Holleran, PhD 44, Avram J Holmes, PhD 75,76,77, Georg Homuth, PhD 78, Jouke-Jan Hottenga, PhD 11,21,22, Hilleke E. Hulshoff Pol, PhD 25, Masashi Ikeda, PhD, MD 79, Neda Jahanshad, PhD 31, Christiane Jockwitz, PhD 14,80, Stefan Johansson, PhD 81,82, Erik G Jönsson, PhD, MD 1,83, Niklas R Jørgensen, PhD, MD, DMSc 84,85, Masataka Kikuchi, PhD 86, Emma EM Knowles, PhD 59,61, Kuldeep Kumar, PhD 5, Stephanie Le Hellard, PhD 87,88, Costin Leu, PhD 89,90,58, David E Linden, PhD, MD 4,42, Jingyu Liu, PhD 28, Arvid Lundervold, PhD, MD 64,91, Astri Johansen Lundervold, PhD 92, Anne M Maillard, PhD 93, Nicholas G Martin, PhD 94, Sandra Martin-Brevet, PhD 95,96, Karen A Mather, PhD 23,97, Samuel R Mathias, PhD 59,61, Katie L McMahon, PhD 98, Allan F McRae, PhD 99,100, Sarah E Medland, PhD 101, Andreas Meyer-Lindenberg, MD 102, Torgeir Moberget, PhD 1, Claudia Modenato,

MSc 103,104, Jennifer Monereo Sánchez, MSc 105,106, Derek W Morris, PhD 44, Thomas W Mühleisen, PhD 14,15,32, Robin M Murray, MD 107, Jacob Nielsen, PhD 49, Jan E Nordvik, PhD 108, Lars Nyberg, PhD 16,17,109, Loes M Olde Loohuis, PhD 110, Roel A Ophoff, PhD 110, Michael J Owen, PhD, FRCPsych 42, Tomas Paus, PhD, MD 111,112, Zdenka Pausova, MD 19,112, Juan M Peralta, MSc 20, Bruce Pike, PhD 113, Carlos Prieto, PhD 114, Erin B Quinlan, PhD 115, Céline S Reinbold, PhD 32,33,48, Tiago Reis Marques, PhD, MD 116,117, James JH Rucker, PhD, MD 118, Perminder S Sachdev, PhD, MD 23,119, Sigrid B Sando, PhD, MD 120,66, Peter R Schofield, PhD, DSc 121,122, Andrew J. Schork, PhD 123,124, Gunter Schumann, MD PhD 115, Jean Shin, PhD 19,112, Elena Shumskaya, PhD 50,125, Ana I Silva, MSc 43,126, Sanjay M Sisodiya, PhD, FRCP 89, Vidar M Steen, PhD, MD 87,88, Dan J Stein, PhD, MD 38,127, Lachlan T Strike, PhD 128, Ikuo K. Suzuki, PhD 129,130,131, Christian K Tamnes, PhD 1,132,133, Alexander Teumer, PhD 134, Anbupalam Thalamuthu, PhD 23, Diana Tordesillas-Gutiérrez, PhD, MSc 35,135, Anne Uhlmann, PhD 38, Magnús Ö Úlfarsson, PhD 7,13, Dennis van 't Ent, PhD 11,21, Marianne BM van den Bree, PhD 42,68, Pierre Vanderhaeghen, Phd 137,138,139, Evangelos Vassos, PhD, MD 107,140, Wei Wen, PhD 23, Katharina Wittfeld, PhD 62,63, Margaret J Wright, PhD 100,141, Ingrid Agartz, PhD 1,83,133, Srdjan Djurovic, PhD 87,142, Lars T Westlye, PhD, MD 1,3, 48, Hreinn Stefánsson, PhD 7, Kári Stefánsson, PhD 7,8, Sébastien Jacquemont, MD 5,143, Paul M Thompson, PhD 31, Ole A Andreassen, PhD, MD 1, for the ENIGMA-CNV working group,

Affiliations:

1. Department of Medical Genetics, Oslo University Hospital, Oslo, Norway.
2. Norwegian Centre for Mental Disorders Research, Division of Mental Health and Addiction, Oslo University Hospital and Institute of Clinical Medicine, University of Oslo, Oslo, Norway
3. KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway.
4. School of Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, the Netherlands
5. Sainte Justine Hospital Research Center, Montreal, Quebec, Canada
6. Centre de recherche de l'Institut universitaire de gériatrie de Montréal, Montreal, Quebec, Canada
7. deCODE Genetics (Amgen), Reykjavík, Iceland
8. Faculty of Medicine, University of Iceland, Reykjavík, Iceland
9. Department of Clinical Biochemistry, Copenhagen University Hospital, Rigshospitalet, Glostrup, Denmark
10. Department of Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, the Netherlands

11. Department of Biological Psychology and Netherlands Twin Register, VU University Amsterdam, Amsterdam, the Netherlands
12. University of Melbourne Academic Unit for Psychiatry of Old Age, Kew, Australia
13. National Ageing Research Institute, Parkville, Australia
14. Institute of Neuroscience and Medicine, INM-1, Research Centre Juelich, Juelich, Germany
15. C. and O. Vogt Institute for Brain Research, Medical Faculty, Heinrich Heine University Duesseldorf, Duesseldorf, Germany
16. Umeå Centre for Functional Brain Imaging, Umeå University, Umeå, Sweden
17. Department of Integrative Medical Biology, Umeå University, Umeå, Sweden
18. Mathematics and Statistics, Murdoch University, Perth, Australia
19. Research Institute, Hospital for Sick Children, Toronto, Ontario, Canada
20. South Texas Diabetes and Obesity Institute, Department of Human Genetics, School of Medicine, University of Texas Rio Grande Valley, Brownsville
21. Amsterdam Neuroscience, Amsterdam, the Netherlands
22. Amsterdam Public Health Research Institute, VU Medical Center, Amsterdam, the Netherlands
23. Centre for Healthy Brain Ageing, School of Psychiatry, University of New South Wales, Sydney, Australia
24. Dementia Centre for Research Collaboration, School of Psychiatry, University of New South Wales, Sydney, Australia
25. Department of Psychiatry, University Medical Center Brain Center, Utrecht University, Utrecht, the Netherlands
26. Institute of Diagnostic Radiology and Neuroradiology, University Medicine Greifswald, Greifswald, Germany
27. Altrecht Science, Utrecht, the Netherlands
28. Tri-institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta
29. The Department of Electrical and Computer Engineering, University of New Mexico, Albuquerque
30. Institute for Anatomy I, Medical Faculty, Heinrich Heine University D
31. Imaging Genetics Center, Mark and Mary Stevens Institute for Neuroimaging and Informatics, University of Southern California, Los Angeles
32. Department of Biomedicine, University of Basel, Basel, Switzerland
33. Institute of Medical Genetics and Pathology, University Hospital Basel, Basel, Switzerland
34. Department of Psychosis Studies, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
35. University Hospital Marqués de Valdecilla, IDIVAL, Centro de Investigación Biomédica en Red Salud Mental (CIBERSAM), Santander, Spain

36. University Hospital Virgen del Rocío, IBiS, Centre de Investigació Biomédica en Red Salud Mental (CIBERSAM), Sevilla, Spain
37. Center for Multimodal Imaging and Genetics, University of California, San Diego
38. Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, Western Cape, South Africa
39. Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
40. Faculty of Health and Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Australia
41. Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
42. MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, United Kingdom
43. Cardiff University Brain Research Imaging Centre School of Psychology, Cardiff University, Cardiff, United Kingdom
44. Centre for Neuroimaging and Cognitive Genomics, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Galway, Ireland
45. Laboratory for Research in Neuroimaging, Department of Clinical Neurosciences, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland
46. Psychological and Social Medicine, Faculty of Medicine, Dresden University of Technology, Dresden, Germany
47. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands
48. Department of Psychology, University of Oslo, Oslo, Norway
49. Signal Transduction, H. Lundbeck A/S, Ottiliavej 9, DK-2500 Valby, Denmark
50. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands
51. Department of Neurology, Akershus University Hospital, 1474 Nordbyhagen, Norway
52. Institute of Clinical Medicine, Campus Ahus, University of Oslo, Oslo, Norway
53. Neurospin, Le Commissariat à l'énergie atomique et aux énergies alternatives, Université Paris-Saclay, Gif-sur-Yvette, France
54. Division of Cerebral Integration, National Institute for Physiological Sciences, Okazaki, Japan
55. Department of Life Science, Sokendai, Hayama, Japan
56. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
57. Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA
58. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
59. Boston Children's Hospital, Boston, Massachusetts

60. Institute of Living, Hartford, Connecticut

61. Harvard Medical School, Boston, Massachusetts

62. Department of Psychiatry and Psychotherapy, Greifswald, Germany

63. German Center of Neurodegenerative Diseases (DZNE), Rostock/Greifswald, Greifswald, Germany

64. Department of Biomedicine, University of Bergen, Bergen, Norway

65. Division of Psychiatry, Haukeland University Hospital, Bergen, Norway

66. Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

67. St Olav's Hospital, Department of Radiology and Nuclear Medicine, Trondheim, Norway

68. School of Medicine, Cardiff University, Cardiff, United Kingdom

69. Department of Pathology of Mental Diseases, National Institute of Mental Health, National Center of Neurology and Psychiatry, Kodaira, Japan

70. Osaka University, Osaka, Japan

71. Princess Màxima Center for Pediatric Oncology, Utrecht, the Netherlands

72. Genentech, Inc., South San Francisco, CA, USA 94080

73. Department of Child and Adolescent Psychiatry/Psychology, Erasmus MC-Sophia's Children's Hospital, Rotterdam, the Netherlands

74. Institute of Human Genetics, University of Bonn Medical School, Bonn, Germany

75. Psychology Department, Yale University, New Haven, CT, USA

76. Department of Psychiatry, Yale University, New Haven, CT, USA

77. Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

78. University Medicine Greifswald, Greifswald, Germany

79. Department of Psychiatry, Fujita Health University School of Medicine, Toyoake, Japan

80. Department of Psychiatry, Psychotherapy and Psychosomatics, RWTH Aachen University, Medical Faculty, Aachen, Germany

81. Department of Clinical Science, University of Bergen, Bergen, Norway

82. Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway

83. Centre for Psychiatric Research, Department of Clinical Neuroscience, Karolinska Institutet, & Stockholm Health Care Services, Stockholm County Council

84. Department of Clinical Biochemistry, Copenhagen University Hospital Rigshospitalet, Glostrup, Denmark

85. OPEN, Odense Patient data Explorative Network, Odense University Hospital/Institute of Clinical Research, University of Southern Denmark, Odense, Denmark

86. Department of Genome Informatics, Graduate School of Medicine, Osaka University, Osaka, Japan

87. Norwegian Centre for Mental Disorders Research, Department of Clinical Science, University of Bergen, Bergen, Norway
88. Dr Einar Martens Research Group for Biological Psychiatry, Department of Medical Genetics, Haukeland University Hospital, Bergen, Norway
89. Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology and Chalfont Centre for Epilepsy, London, United Kingdom
90. Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, United States
91. Mohn Medical Imaging and Visualization Centre, Department of Radiology, Haukeland University Hospital, Bergen, Norway
92. Department of Biological and Medical Psychology, University of Bergen, Bergen, Norway
93. Service des Troubles du Spectre de l'Autisme et apparentés, Lausanne University Hospital, Switzerland
94. Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Australia
95. Service of Medical Genetics, Lausanne University Hospital, Lausanne, Switzerland
96. LREN, Department of clinical neurosciences, Lausanne University Hospital, Lausanne, Switzerland
97. Neuroscience Research Australia, Randwick, Australia
98. Herston Imaging Research Facility and School of Clinical Sciences, Queensland University of Technology, Brisbane, Australia
99. Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
100. Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia
101. Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Australia
102. Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, University of Heidelberg, Mannheim, Germany
103. LREN - CHUV, Lausanne, Switzerland
104. University of Lausanne, Lausanne, Switzerland
105. Department of Radiology and Nuclear Medicine, Maastricht University Medical Center, Maastricht, Netherlands
106. School for Mental Health and Neuroscience, Maastricht, Netherlands
107. Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
108. The CatoSenteret Rehabilitation Center, Son, Norway
109. Department of Radiation Sciences, Umeå University, Umeå, Sweden
110. Center for Neurobehavioral Genetics, University of California, Los Angeles
111. Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario, Canada
112. Physiology and Nutritional Sciences, University of Toronto, Toronto, Ontario, Canada
113. Department of Radiology, University of Calgary, Calgary, Alberta, Canada
114. Bioinformatics Service, Nucleus, University of Salamanca, Salamanca, Spain

115. Centre for Population Neuroscience and Precision Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
116. Department of Psychosis, Institute of Psychiatry, Psychology & Neuroscience, Kings College, London, United Kingdom
117. Psychiatric Imaging Group, MRC London Institute of Medical Sciences (LMS), Hammersmith Hospital, Imperial College, London, United Kingdom
118. Institute of Psychiatry, Psychology and Neuroscience, London, London, United Kingdom
119. Neuropsychiatric Institute, The Prince of Wales Hospital, Sydney, Australia
120. Department of Neurology, University Hospital of Trondheim, Trondheim, Norway
121. Neuroscience Research Australia, Sydney, Australia
122. School of Medical Sciences, University of New South Wales, Sydney, Australia
123. Institute of Biological Psychiatry, Roskilde, Denmark
124. The Translational Genetics Institute (TGEN), Phoenix, AZ, United States
125. Department of Human Genetics, Radboud University Medical Center, Nijmegen, the Netherlands
126. Neuroscience and Mental Health Research Institute, Cardiff, United
127. South African Medical Research Council Unit on Risk and Resilience in Mental Disorders, Department of Psychiatry and Neuroscience Institute, University of Cape Town, Cape Town, South Africa
128. Queensland Brain Institute, The University of Queensland, Brisbane, Australia
129. VIB Center for Brain & Disease Research, Stem Cell and Developmental Neurobiology Lab, Leuven, Belgium
130. University of Brussels (ULB), Institute of Interdisciplinary Research (IRIBHM) ULB Neuroscience Institute, Brussels, Belgium
131. The University of Tokyo, Department of Biological Sciences, Graduate School of Science, Tokyo, Japan
132. PROMENTA Research Center, Department of Psychology, University of Oslo, Oslo, Norway
133. Department of Psychiatry, Diakonhjemmet Hospital, Oslo, Norway
134. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
135. Neuroimaging Unit, Technological Facilities, Valdecilla Biomedical Research Institute, IDIVAL, Santander, Spain
136. Faculty of Electrical and Computer Engineering, University of Iceland, Reykjavík, Iceland
137. VIB-KU Leuven Center for Brain & Disease Research, 3000 Leuven, Belgium
138. KU Leuven, Department of Neurosciences & Leuven Brain Institute, 3000 Leuven, Belgium
139. Université Libre de Bruxelles (U.L.B.), Institut de Recherches en Biologie Humaine et Moléculaire (IRIBHM), and ULB Neuroscience Institute (UNI), 1070 Brussels, Belgium
140. National Institute for Health Research, Mental Health Biomedical Research Centre, South London and Maudsley National Health Service Foundation Trust and King's College London, London, United Kingdom

141. Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia

142. Department of Medical Genetics, Oslo University Hospital, Oslo, Norway

143. Department of Pediatrics, University of Montreal, Montreal, Quebec, Canada

* Corresponding author: Ida Elken Sønderby. E-mail: i.e.sonderby@medisin.uio.no.
ORCID-ID: https://orcid.org/0000-0001-7297-7855.
Postal address: Department of Medical Genetics, Oslo University Hospital, Postboks 4956 Nydalen, 0424 Oslo,, Norway

Keywords: copy number variant; 1q21.1 distal; neuroimaging, neurodevelopment, cortical surface area,

Word count, abstract/body: 209/3,889.

Tables/figures/supplements: *3/2/2* (supplementary tables & supplementary figures and info)

Running title: 1q21.1 distal CNVs influence cortical surface area

Abstract

Low-frequency 1q21.1 distal deletion and duplication copy number variant (CNV) carriers are predisposed to multiple neurodevelopmental disorders including schizophrenia, autism and intellectual disability. Human carriers display a high prevalence of micro- and macrocephaly in deletion and duplication carriers, respectively. The underlying brain structural diversity remains largely unknown.

We systematically called CNVs in 38 cohorts from the large-scale ENIGMA-CNV collaboration and the UK biobank and identified 28 1q21.1 distal deletion and 22 duplication carriers and 37,088 non-carriers (48 % male) derived from 15 distinct MRI scanner sites. With standardized methods, we compared subcortical and cortical brain measures (all) and cognitive performance (UK biobank only) between carrier groups also testing for mediation of brain structure on cognition. We identified positive dosage effects of copy number on intracranial volume (ICV) and total cortical surface area, with largest effects in frontal and cingulate cortices, and negative dosage effects on caudate and hippocampal volumes. The carriers displayed distinct cognitive deficit profiles in cognitive tasks from the UK biobank with intermediate decreases in duplication carriers and somewhat larger in deletion carriers – the latter potentially mediated by ICV or cortical surface area. These results shed light on pathobiological mechanisms of neurodevelopmental disorders, by demonstrating gene dose effect on specific brain structures and effect on cognitive function.

Introduction

Inter-individual differences in brain structure are highly heritable[1], but identifying the genes that contribute to brain development is challenging. Genome-wide association studies (GWAS) of brain anatomical structures indicate the influence of many single

nucleotide polymorphisms (SNPs) with small effect sizes[2, 3], but the links to brain function remain weak. Evidence is emerging that some rare copy number variants (CNVs) - i.e., regions of the genome that are either deleted or duplicated - are associated with both substantial brain size and shape differences; e.g., the 7q11.23[4, 5], 22q11.2[6, 7], 15q11.2[8-11] and 16p11.2 proximal[12-14] and distal CNVs[15]. Many of these CNVs also have wide-ranging phenotypic impact, including poorer cognitive abilities[8, 16-18] and increased risk of neurological or neurodevelopmental disorders. The strong impact of these CNVs on brain structure and behavior make them valuable for studies of the molecular mechanisms contributing to aberrant human neurodevelopment.

The 1q21.1 distal CNV has a known large effect on head circumference, as evident from a high prevalence of micro- and macrocephaly in deletion and duplication carriers, respectively[19-21]. This, along with its position in a region that is rich in genes unique to the human lineage (i.e.absent in primates)[22, 23], makes the 1q21.1 distal CNV particularly interesting for the study of aberrations in human brain structure. However, its relatively low frequency, 1 in ~3,400, (deletions) and 1 in 2,100 (duplications)[8, 16], has hampered the study of its effects on brain structure.

1q21.1 distal deletion and duplication carriers are both at higher risk for several neurodevelopmental disorders including schizophrenia[24], intellectual disability (ID), developmental delay (DD), speech problems, autism spectrum disorders (ASD), motor impairment[19, 25-27] and epilepsy[25, 28], in addition to separate risk for the duplication carriers for ADHD[29], bipolar disorder and major depression[30, 31]. Further, general cognitive ability (IQ) was lower in carriers in a small clinical study[19] and in the UK biobank[32]. In addition, 1q21.1 distal CNVs display a positive dose response on head circumference[19-21], height and weight[33, 34] and are associated with various somatic

297 diseases and traits including bone and muscle deviations[33] and cataract[35] (deletion only),
298 diabetes[35] (duplication only) and heart disease[35-38] (both). Conversely, several studies
299 report carriers without any clinically evident phenotypes[19, 37] and considerable
300 heterogeneity[39, 40], suggesting incomplete penetrance and variable expressivity. The
301 Df(h1q21) +/- mouse, deleted in the syntenic 1q21.1 distal region, displays some
302 phenotypes similar to human CNV carriers, including reduced head-to-tail-length and
303 altered dopamine transmission in response to psychostimulants, as seen in people with
304 schizophrenia[41].

305 The 1q21 region in humans is rich in low copy number repeats[20, 42] and contains several
306 recurrent CNVs with differing breakpoints[21, 36]. Thus, gene estimates vary, but the core
307 interval encompasses at least twelve protein-coding genes including several human-
308 specific genes such as *HYDIN2*[21, 36], *NOTCH2NLs*[22, 23] and the DUF1220/Olduvain
309 domain containing *NBPF*-encoding genes[43-45] – the two latter were recently shown to
310 have evolved as a two-gene unit[46]. Particularly interesting in the context of brain
311 development are the recently characterized *NOTCH2NL* genes, absent in human's closest
312 living relatives and shown to prolong cortical neurogenesis[22, 23].
313
314 Despite the strong effects on neurodevelopmental traits and disorders, the impact of the
315 1q21.1 CNVs on human brain structure is largely unknown. Here, we present the first
316 large-scale systematic neuroimaging study of 1q21.1 distal CNV carriers, investigating
317 brain structure in more than 37,000 individuals including 28 deletion and 22 duplication
318 carriers. We mapped the effect of the 1q21.1 distal CNV on subcortical volumes,
319 intracranial volume (ICV) and global and regional measures of mean cortical thickness
320 and surface area. We investigated variation in cognitive task performance and
321 supplemented with exploratory mediation analysis of the brain on cognition in the UK

Biobank. Given prior findings[19-21, 47], we explored a dose-dependent effect of copy number on brain structures and decreased cognitive performance for both 1q21.1 distal deletion and duplication carriers in comparison to non-carriers.

Materials and Methods

Sample description

The brain structural sample comprises a total of 39 cohorts with genotyping and MRI imaging data – 38 from the ENIGMA-CNV consortium in addition to a subsample of the UK Biobank[48] (project ID number #27412). Demographic characteristics for each cohort is described in Supplementary Table 1 with a reference to participants' collection and data sets including individual inclusion and exclusion parameters. Extended information on diagnosis and family information can be found in Supplementary Note 1 and age distribution of the cohorts in Supplementary Figure 1. All participants gave written informed consent and sites involved obtained ethical approvals. The main 1q21.1 distal sample consisted of 28 deletion carriers, 22 duplication carriers and 37,088 non-carriers (Table 1) from 13 different datasets and 15 scanner sites with various ascertainments (family, clinical and population studies, case-control study for psychiatric disease) collected up until Sep 30, 2019. Non-carriers were defined as having no CNVs known to cause neurodevelopmental diseases (as defined in Supplementary Table 2). In the meta-analysis, an independent Icelandic sample from deCODE Genetics consisting of two deletion carriers and five duplication carriers in addition to 1150 non-carriers was added.

Genotyping and QC

The genotypes were obtained by genotyping with commercially available platforms, performed at participating sites for each cohort (Supplementary Table 1). Individuals were excluded exclusively based on quality control (QC) parameters from the CNV

calling. No exclusion was done due to ancestry in the primary analysis but the effect of ancestry was evaluated in a separate analysis (see below).

CNV calls and validation in the core ENIGMA-CNV sample

Almost all cohorts had CNVs called and identified in a unified manner as described previously[15]. In brief, CNVs were called using PennCNV[49] and appropriate population frequency (PFB)-files and GC (content)-model files (Supplementary Table 3) (See Supplementary Notes 2 and 3). Samples were filtered and CNVs identified based on standardized quality control metrics[15] (Supplementary Notes 2 and 3). The 1q21.1 distal region was well-covered by all arrays (Supplementary Figure 2). CNVs overlapping the region of interest (1q21.1 distal and 1q21.1 distal and proximal) were identified with the R package iPsychCNV, visualized and manually inspected.

Image acquisition and processing

All brain measures were obtained from structural T1-weighted MRI data collected at participating sites around the world and analyzed with the standardized image analysis, FreeSurfer, quality assurance and statistical methods as per the harmonized neuroimaging protocols developed within ENIGMA2[3] and ENIGMA3 (http://enigma.ini.usc.edu/protocols/imaging-protocols/). Further detail on data processing is provided in Supplementary Note 4. Details on study, scanner, vendor, field strength, sequence, acquisition parameters and FreeSurfer versions used are outlined in Supplementary Table 4.

Statistical Analysis

Imaging data processing and CNV calling were performed locally and de-identified CNV and imaging data were provided for a central mega-analysis. One of a pair of duplicates was kept. Relatives were removed from the sample used for the main analysis. In

addition, we conducted a number of sensitivity analyses to test the robustness of the results (Supplementary Note 5, Supplementary Tables 5, 6, 7 and 8). Individuals with a minimum overlap of 0.4 to regions with known pathogenic CNVs (Supplementary Table 2) were excluded from the analysis regardless of copy number status as were individuals from scanner sites without 1q21.1 distal CNV carriers.

Brain measures were normalized in R v3.3.2 by an inverse normal transformation of the residual of a linear regression on the phenotype correcting for covariates as done previously[15]. For the primary analysis, covariates were age, age^2, sex, scanner site and ICV. In the analysis of ICV, ICV was not included as a covariate. These final covariance-corrected values were used in downstream analysis and are reported for each measure. For comparison between groups, normalization was carried out including only the groups addressed (deletion and non-carriers, duplications and non-carriers) except for the deletion versus duplication comparison, where values from normalization of the entire dataset was used due to the low numbers.

For the copy number dosage effect analysis (i.e. the effect on brain structure of 1q21.1 distal copy number variation), a linear regression on the copy number status of the individuals (deletion=1, normal=2, duplication=3) was performed using the following model: covariance-corrected, normalized brain measure ~ copy number (deletion=1, non-carrier=2, duplication=3). For comparison between groups, a two sample two-sided t-test assuming equal variance in all carrier/non-carrier groups was employed (R v3.3.2) where deletion or duplication carriers were compared either to each other or to non-carriers. To correct for the multiple comparisons, we calculated the number of independent outcome measures through spectral decomposition of a correlation matrix using MatSpDlite

(https://neurogenetics.qimrberghofer.edu.au/matSpDlite/) of the three global, seven subcortical and 68 regional cortical measures. Based on the ratio of observed eigenvalue variance to its theoretical maximum, the estimated equivalent of independent measures was 36. Thus, we set the significance threshold at $\alpha=0.05/36=0.0014$. We report the uncorrected p-values throughout the manuscript.

Effect size is calculated as the absolute effect size (the difference in mean between the two copy number groups in the t-test – which, in this case, equals Cohen's *d* as the standard deviation of the normalized brain measures is one) and the estimate of beta in the linear regression. Plots were generated using R library ggplot2 v2.2.1[50]. Regional cortical visualisation was done with the R package ggseg v1.5.1.

In a novel analysis, the independent Icelandic data was processed and analyzed as the main dataset. We meta-analyzed the results using the R package *metafor* v2.0.0, as previously[15].

Cognitive task performance data

We downloaded behavioral performance measures on seven cognitive tests (the pairs matching task, the reaction time task, reasoning and problem solving tests, the digit span test, the symbol digit substitution test and the trail making A and B tests) from the UK Biobank repository, performed by at least 10% of the participants. The results were processed following the general approach by Kendall et al[16]. For more details, see Supplementary Note 6. For the analysis of the seven cognitive measures, we set the significance threshold to $\alpha=.05/7=.007$.

Mediation analysis

Mediation analyses were done with the R package *mediation* v4.4.7. Brain measures were normalized as described above and cognitive tasks were corrected for age, age^2 and sex prior to input into the analysis. We report the proportion of the total effect of the CNV on cognitive task performance mediated by the brain measures ("path ab"/"path c"), with p-values calculated through quasi-Bayesian approximation using 5000 simulations. We set the significance threshold at $\alpha=.05/((2+4)x6)=1.4 \times 10^{-3}$ given the test of two structures for deletion and four for duplication carriers on six cognitive tests. The digit span test was excluded since no 1q21.1 CNV carriers had results from both this cognitive test and brain structural data.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information files. The data was gathered from various resources, and material requests will need to be placed with individual PIs. I.E.S. can provide additional detail upon correspondence. Data from PING is available at NIMH Data Archive: https://ndar.nih.gov/edit_collection.html?id=2607

Results

Sample characteristics

The main 1q21.1 distal (146.5-147.4Mb, hg19) brain structural dataset consisted of 28 deletion and 22 duplication carriers and 37,088 non-carriers (derived from the same scanner sites as the CNV carriers) from ENIGMA-CNV and UK biobank (Table 1, separate demographics in Supplementary Table 9). The age of CNV carriers was lower (41.7±19.0 (deletions), 55.4±12.7 (duplications), respectively) than that of non-carriers (61.1±12.1) (Table 1). Eleven deletion carriers and seven duplication carriers had a

446 known neurological, neurodevelopmental or psychiatric diagnosis or had been recruited
447 in a clinical CNV study. The remaining carriers either did not have an established
448 diagnosis or were recruited in studies from which diagnostic information was unavailable
449 (Table 1, Supplementary Table 10). Of the 37,088 non-carriers, 6.5 % (2,425) had an
450 established neurological, neurodevelopmental or psychiatric disorder.
451
452 **1q21.1 distal CNV associated with global cortical surface structures**
453 For our main dataset, there was a significant positive association between the number of
454 1q21.1 distal copies and ICV ($\beta = 1.47$, $P = 2.8 \times 10^{-25}$) as well as cortical surface area (β
455 $= 0.81$, $P = 1.1 \times 10^{-8}$) (Figure 1, Supplementary Table 5) at a significance threshold of
456 $P<0.0014$ after correction for age, age^2, sex, scanner site and ICV. In contrast, a
457 significant negative copy number dosage effect was identified for the caudate ($\beta = -0.49$,
458 $P = 6.9 \times 10^{-4}$) and hippocampal volumes ($\beta = -0.56$, $P = 1.3 \times 10^{-4}$) T-tests indicated a
459 decrease in ICV (Cohen's D = -1.84 (-17%), $P = 1.6 \times 10^{-22}$) for deletion carriers and an
460 increase for duplication carriers (Cohen's D = 0.90 (+10%), $P = 2.3 \times 10^{-5}$), respectively,
461 compared to non-carriers (Supplementary Table 6). For a raw value plot of ICV, see
462 Supplementary Figure 3. The cortical surface area dosage effect was primarily driven by
463 the deletion carriers with a significantly lower total cortical surface area (Cohen's D = -
464 1.13 (-23%), $P = 2.1 \times 10^{-9}$) and the dosage effect on caudate and hippocampus was
465 primarily driven by duplication carriers with significantly smaller caudate (Cohen's D = -
466 0.71 (-16%), P = 0.0012) and hippocampal (Cohen's D = -0.92 (-15%), $P = 4.1 \times 10^{-5}$)
467 volumes than non-carriers (Figure 1, Supplementary Table 7). Adding an independent
468 Icelandic dataset with two deletion, five duplication and 1150 non-carriers (Table 1) in a
469 meta-analysis strengthened the majority of the dosage results (Supplementary Figure 4,

Supplementary Tables 11 and 12) and revealed additional significant between-group differences in nucleus accumbens, caudate and putamen (Supplementary Table 12).

A number of sensitivity analyses was run on the main dataset, namely:

a) Matching each carrier with one non-carrier for age, sex, scanner site and ICV or age, sex, scanner site;
b) including only: i) non-affected individuals (i.e. excluding individuals with a known neurodevelopmental or neurological disorder diagnosis; ii) adults (age>=18); iii) non-affected adults; iv) children (age<18); v) ENIGMA-CNV or vi) UK biobank;
c) controlling for ancestry;
d) excluding ICV as a covariate or;
e) including first and second-degree relatives (see Supplementary Note 5 for methods).

These analyses validated the overall effects (Supplementary Tables 5 and 6).

The 1q21.1 distal CNV is associated with regional brain structures

The largest dosage effects for regional cortical surface area were found in the frontal lobes followed by the cingulate cortex - with additional significant effects in three regions of the parietal and temporal lobes (Figure 2, Supplementary Table 7). Likewise, through *t*-tests, the largest effects in both deletion and duplication carriers in comparison to non-carriers were observed in the frontal and cingulate cortices (Figure 3, , Supplementary Table 8).

For regional cortical mean thickness, we identified significant negative dosage effects in the superior temporal region and significant positive dosage effects for the pericalcarine region (Figure 2, Supplementary Tables 7 and 8). Similarly, significant increases in mean

cortical thickness were observed in deletion carriers versus non-carriers in the parstriangularis and superior temporal regions and a significant decrease in the pericalcarine region (Figure 2, Supplementary Table 8). All regional results were corrected for age, age^2, sex, scanner site and ICV. Sensitivity analyses similar to those performed for subcortical regions confirmed the robustness of the results (Supplementary Tables 7 and 8).

1q21.1 distal CNV associated with cognitive performance and mediation by brain structures

Deletion and duplication carriers had different cognitive profiles in comparison to non-carriers when testing for association in seven different neuropsychological tests available from the full UK biobank sample: Deletion carriers had significantly poorer performance in three tests: symbol digit substitution, trail making B and pairs matching while duplication carriers had significantly poorer performance in two tests: reaction time and the reasoning and problem solving task (Table 2).

Testing the effect of brain structures on cognitive tests in UK biobank participants, larger ICV and total surface area were associated with better performance on almost all tests (Table 3, see Supplementary Table 13 for sample size details). A larger hippocampus was associated with better performance for symbol digit substitution, trail making A and B (Table 3) and a larger caudate was associated with higher performance on trail making A (Table 3).

Next, we tested whether the brain structures significantly associated with 1q21.1 distal CNV carriers might mediate the effect of the CNV on cognition. For two of the three tests associated with deletion carrier status, there were significant mediation effects (significance threshold 1.4×10^{-3}): Cortical surface area and ICV accounted for 5% and

520 10%, respectively, of the poorer performance of deletion carriers on symbol digit
521 substitution, and 7% and 17%, respectively, of their poorer performance on the trail
522 making B test (Table 3).
523
524 **Discussion**
525 Our main finding was a significant positive dosage effect in humans of 1q21.1 distal copy
526 number on ICV and cortical surface area, with the largest differences in frontal and
527 cingulate cortical surface area. We also identified a significant negative dosage effect on
528 caudate and hippocampal volumes. A number of sensitivity analyses confirmed the
529 robustness of the results. Both 1q21.1 distal deletion and duplication carriers showed
530 poorer cognitive performance, although on different tests, with an indication that
531 decreased ICV/cortical surface area might mediate the effect in deletion carriers.

532 ***The 1q21.1 distal CNV causes copy dosage effect on brain structures.***
533 We found a strong effect of the 1q21.1 distal CNV on total cortical surface area while no
534 overall effect on mean cortical thickness was observed. A specific increase in size of the
535 cortical surface area with little effect on cortical thickness is observed throughout
536 mammalian evolution including the primate lineage leading to humans[51]. This possibly
537 reflects that cortical thickness and surface area appear to be driven by distinct genetic
538 processes[52]. This pattern may be the result of an increased number of symmetric or self-
539 renewing cell-division cycles, leading to an expansion of the neural progenitor pool and
540 subsequently to an increase in the number of cortical neurons - in line with the radial unit
541 hypothesis[51]. Interestingly, although not significant, mean cortical thickness tended to
542 decrease in deletion carriers in the frontal cortical surface areas with the highest effect
543 sizes, resembling a pattern found in lissencephaly[53]. This could suggest that large

544 regional decreases in cortical surface area correlate inversely with mean cortical
545 thickness.
546 The biomechanical forces of brain growth are thought to form the expansion of the
547 cranium so that the skull grows in harmony with the expanding brain[54]. Thus, the positive
548 copy number dosage effect on cortical surface area may directly trigger the effect on head
549 circumference[19-21] and ICV of 1q21.1 distal carriers due to modifications in pressure.
550 Altered mechanical pressure might also cause the negative copy number dosage effect on
551 hippocampus and caudate volumes, effects on subcortical volumes also observed in a UK
552 biobank exploratory study on six individuals with a 1q21.1 distal duplication [55].
553
554 ***Human-specific genes may affect the cortical surface area*** **and cross-species effects**
555 The positive copy number dosage effect on brain structure with the same direction as for
556 weight and height[33, 34] likely results from altered gene expression as observed in 1q21.1
557 distal CNV cell lines[47]. In an independent experiment on fetal tissue, we also observed a
558 dynamic expression patterns of the genes in the 1q21.1 interval consistent with potential
559 roles in cortical neurogenesis and development (Supplementary Note 7, Supplementary
560 Figures 5 & 6). Genome-wide association studies (GWAS) based on the hg19 genome
561 assembly have not identified hits in the 1q21.1 genomic region for ICV[56], total cortical or
562 regional surface area[52, 57]. Assembly of the 1q21.1 region[58] and thus gene discovery is
563 complicated due to the presence of numerous low copy number repeats[20, 42] and has been
564 faulty until the GRCh38 genome assembly. This may explain the lack of GWAS hits in
565 the region.
566 Candidates for a dosage-dependent amplifier of the CNV-associated brain phenotypes are
567 the recently identified human-specific NOTCH2NL genes that confer delayed neuronal
568 differentiation and increased progenitor self-renewal[22, 23] - in line with the radial unit

hypothesis[51]. The areas with the highest regional effect sizes overlap with the areas of the highest expression of NOTCH2NLA and C in utero[22] in concordancw ith n early developmental effect such as the macrocephaly observed *in utero* in a 1q21.1 distal duplication carrier[37]. Our observations of a 2 % reduced skull diameter in the 1q21.1 deletion mouse (Supplementary Figure 7, Supplementary Notes 8 and 9) and recent findings of decreased total brain volume focused on the temporo-parietal and subcortical areas in the deletion mouse[59], suggests that genes overlapping between human and mice (nine of ten mice genes are syntenic to the human region[41]), and not specific to humans are also involved in the altered skull and brain morphology. However, although diameter and volume are not directly comparable, the 17% decrease in ICV in human 1q21.1 deletion carriers would still point towards a substantial role of human-specific genes or genes with altered functions in comparison to mice. This underlines the need for additional data to disentangle which specific genes are involved in the skull and brain structural phenotypes. Of note, we also observed shorter bones overall in the 1q21.1 deletion mice (Supplementary Figure 7, Supplementary Note 9) expanding on previous head-tail length data [41] and lower bone mineral density in female mice (Supplementary Figure 8, Supplementary Note 9) which mirror bone characteristics from human deletion carriers[33] increasing the number of observed cross-species effects between the 1q21.1 mice and human 1q21.1 deletion carriers.

1q21.1 distal CNV deletion and duplication carriers show deficits in different cognitive functions.

Our findings of widespread lower performance across several tests in different domains for both carrier groups in the volunteer-based UK Biobank sample are in line with cognitive results from a recent study[32] and support that cognitive function in CNV carriers largely without a neurodevelopmental diagnosis may still be compromised [8, 16].

Interestingly, the frontal and cingulate regions[60], with the greatest cortical effect sizes for 1q21.1 distal correlate particularly with cognitive function and have gone through the greatest expansion during human development and evolution [61]. Our analyses indicated that the decreases in cognitive task performance are partially mediated by the observed differences in ICV and cortical surface area, reflecting the positive correlation between brain volume and intellectual function in line with previous findings[62]. The decrease in performance for several cognitive tasks in duplication carriers despite a larger ICV and cortical surface area suggests that the positive correlations may only be applicable within a certain narrower range. Interestingly, recent genetic analysis of NOTCH2NL in archaic and modern humans revealed ongoing adaptive evolution towards a lower dosage of the protein[63] suggesting negative effects of too much NOTCH2NL protein.

Our brain structural findings in 1q21.1 distal CNV carriers overlap with brain alterations in associated disorders: e.g. ADHD[64], ASD[65] schizophrenia[66], bipolar disorder[67], major depressive disorder[68] and subtypes of epilepsy[69] but the exact overlaps differ between carrier groups. Of note, 1q21.1 distal deletion and duplication carriers display direct, opposite effects on several brain structures while at risk for the same neurodevelopmental diseases. Other pathogenic CNVs also display overlapping disease risk and similar opposite copy number effects[6, 8-15] including effects on cortical surface area in 22q11 and 16p11.2 proximal CNV carriers[6, 12-14]. These CNVs impact different genes but may converge on the same downstream pathways altering cortical surface area formation, similar to what has been reported for behavioral and neurocognitive phenotypes[27].

This also suggests that other risk factors interplay to cause disease. It also supports that subgroups within neurodevelopmental disorders can be defined based on genetic profile and brain structural differences.

We demonstrate large effects of 1q21.1 distal CNVs on brain structure and cognition in humans and recapitulation of the smaller head size and other skeletal characteristics in 1q21.1 deletion mice. These findings provide insight into molecular mechanisms involved in critical stages of human brain development and mapping of gene dosages to brain structural fingerprints.

Acknowledgements

1000BRAINS: The 1000BRAINS study was funded by the Institute of Neuroscience and Medicine, Research Center Juelich, Germany. We thank the Heinz Nixdorf Foundation (Germany) for the generous support of the Heinz Nixdorf Recall Study on which 1000BRAINS is based. We also thank the scientists and the study staff of the Heinz Nixdorf Recall Study and 1000BRAINS. Funding was also granted by the Initiative and Networking Fund of the Helmholtz Association (Caspers) and the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 785907 (Human Brain Project SGA2; Amunts, Caspers, Cichon).

Brainscale: The Brainscale study was supported by the Netherlands Organization for Scientific Research MagW 480-04-004 (Boomsma), 51.02.060 (Hilleke Hulshoff Pol), 668.772 (Boomsma & Hulshoff Pol); NWO/SPI 56-464-14192 (Boomsma), the European Research Council (ERC-230374) (Boomsma), High Potential Grant Utrecht University (Hulshoff Pol), NWO Brain and Cognition 433-09-220 (Hulshoff Pol).

Betula: The Betula study was funded by the Knut and Alice Wallenberg (KAW) foundation. The Freesurfer segmentations on the Betula sample was performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at HPC2N in Umeå, Sweden.

Brain Imaging Genetics (BIG): This work makes use of the BIG database, first established in Nijmegen, The Netherlands, in 2007. This resource is now part of Cognomics (www.cognomics.nl), a joint initiative by researchers from the Donders Centre for Cognitive Neuroimaging, the Human Genetics and Cognitive Neuroscience departments of the Radboud university medical centre and the Max Planck Institute for Psycholinguistics in Nijmegen. The Cognomics Initiative has received supported from the participating departments and centres and from external grants, i.e. the Biobanking and Biomolecular Resources Research Infrastructure (Netherlands) (BBMRI-NL), the Hersenstichting Nederland, and the Netherlands Organisation for Scientific Research (NWO). The research leading to these results also receives funding

from the NWO Gravitation grant 'Language in Interaction', the European Community's Seventh Framework Programme (FP7/2007– 2013) under grant agreements n° 602450 (IMAGEMEND), n°278948 (TACTICS), and n°602805 (Aggressotype) as well as from the European Community's Horizon 2020 programme under grant agreement n° 643051 (MiND) and from ERC-2010-AdG 268800-NEUROSCHEMA. In addition, the work was supported by a grant for the ENIGMA Consortium (grant number U54 EB020403) from the BD2K Initiative of a cross-NIH partnership.

deCODE genetics: deCODE genetics acknowledges support from the Innovative Medicines Initiative Joint Undertaking under grant agreements' no. 115008 (NEWMEDS) and no. 115300 (EUAIMS) of which resources are composed of EFPIA in-kind contribution and financial contribution from the European Union's Seventh Framework Programme (EU-FP7/2007-2013), EU-FP7 funded grant no. 602450 (IMAGEMEND) and EU funded FP7-People-2011-IAPP grant agreement no. 286213 (PsychDPC).

Dublin: Work was supported by Science Foundation Ireland (SFI grant 12/IP/1359 to Gary Donohoe and SFI08/IN.1/B1916-Corvin to Aidan C Corvin)

ECHO-DEFINE: The ECHO study acknowledges funding from a Medical Research Council (MRC) Centre Grant to Owen (G0801418), the Wellcome Trust (Institutional Strategic Support Fund (ISSF) to van den Bree and Clinical Research Training Fellowship to Doherty), the Waterloo Foundation (WF 918-1234 to van den Bree), the Baily Thomas Charitable Fund (2315/1 to van den Bree), National Institute of Mental Health (NIMH 5UO1MH101724 to van den Bree and Owen), the IMAGINE-ID study (funded by MRC (MR/N022572/1 to Hall, van den Bree and Owen). The DEFINE study was supported by a Wellcome Trust Strategic Award (100202/Z/12/Z) to Owen.

ENIGMA: ENIGMA is supported in part by NIH grants U54 EB20403, R01MH116147, and R56AG058854. NIA T32AG058507; NIH/NIMH 5T32MH073526

EPIGEN-Dublin: The EPIGEN-Dublin cohort was supported by a Science Foundation Ireland Research Frontiers Programme award (08/RFP/GEN1538).

EPIGEN-UK (Sisodiya): The work was partly undertaken at UCLH/UCL, which received a proportion of funding from the UK Department of Health's NIHR Biomedical Research Centres funding scheme. We are grateful to the Wolfson Trust and the Epilepsy Society for supporting the Epilepsy Society MRI scanner.

GOBS: The GOBS study data collection was supported in part by the National Institutes of Health (NIH) grants: R01 MH078143, R01 MH078111 and R01 MH083824 with work conducted in part in facilities constructed under the support of NIH grant C06 RR020547.

GSP: Data were in part provided by the Brain Genomics Superstruct Project (GSP) of Harvard University and Massachusetts General Hospital (MGH) (Principal Investigators: Randy Buckner, Jordan Smoller and Joshua Roffman), with support from the Center for Brain Science Neuroinformatics Research Group, Athinoula A. Martinos Center for Biomedical Imaging, Center for Genomic Medicine, and Stanley Center for Psychiatric Research. Twenty individual investigators at Harvard and MGH generously contributed data to the overall project. We would like to thank Randy Buckner for insightful comments and feedback on this work

HUBIN: The HUBIN study was financed by the Swedish Research Council (K2010-62X-15078-07-2, K2012-61X-15078-09-3, 521-2014-3487
K2015-62X-15077-12-3, 2017-00949), the regional agreement on medical training and clinical research between Stockholm County Council and the Karolinska Institutet.

HUNT: The HUNT Study is a collaboration between HUNT Research Centre (Faculty of Medicine and Movement Sciences, NTNU – Norwegian University of Science and Technology), Nord-Trøndelag County Council, Central Norway Health Authority, and the Norwegian Institute of Public Health. HUNT-MRI was funded by the Liaison Committee between the Central Norway Regional Health Authority and the Norwegian University of Science and Technology, and the Norwegian National Advisory Unit for functional MRI.

IMAGEN: This work received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network based stratification of reinforcement-related disorders) (695313), ERANID (Understanding the Interplay between Cultural, Biological and Subjective Factors in Drug Use Pathways) (PR-ST-0416-10004), BRIDGET (JPND: BRain Imaging, cognition Dementia and next generation GEnomics) (MR/N027558/1), Human Brain Project (HBP SGA 2, 785907),the FP7 projects IMAGEMEND(602450; IMAging GEnetics for MENtal Disorders) and MATRICS (603016), the Innovative Medicine Initiative Project EU-AIMS (115300-2), the Medical Research Council Grant 'c-VEDA' (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the Swedish Research Council FORMAS, the Medical Research Council, the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministeriumfür Bildung und Forschung (BMBF grants 01GS08152;

01EV0711; eMED SysAlc01ZX1311A; Forschungsnetz AERIAL 01EE1406A, 01EE1406B), the Deutsche Forschungsgemeinschaft (DFG grants, SM 80/7-2, SFB 940/2), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1). Further support was provided by grants from: ANR (project AF12-NEUR0008-01 - WM2NA, ANR-12-SAMA-0004), the Eranet Neuron (ANR-18-NEUR00002-01), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the Fondation de l'Avenir (grant AP-RM-17-013), the Fédération pour la Recherche sur le Cerveau ; the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence.

Lifespan: The study is funded by the Research Council of Norway (230345, 288083, 223273).

NCNG: NCNG sample collection was supported by grants from the Bergen Research Foundation and the University of Bergen, the Dr Einar Martens Fund, the Research Council of Norway, to le Hellard, Steen and Espeseth. The Bergen group was supported by grants from the Western Norway Regional Health Authority (Grant 911593 to AL, Grant 911397 and 911687 to AJL).

NTR: The NTR cohort was supported by the Netherlands Organization for Scientific Research (NWO) and The Netherlands Organisation for Health Research and Development (ZonMW) grants 904-61-090, 985-10-002, 912-10-020, 904-61-193,480-04-004, 463-06-001, 451-04-034, 400-05-717, Addiction-31160008, 016-115-035, 481-08-011, 056-32-010, Middelgroot-911-09-032, OCW_NWO Gravity program –024.001.003, NWO-Groot 480-15-001/674, Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI –NL, 184.021.007 and 184.033.111); Spinozapremie (NWO- 56-464-14192), KNAW Academy Professor Award (PAH/6635) and University Research Fellow grant (URF) to DIB; Amsterdam Public Health research institute (former EMGO+) , Neuroscience Amsterdam research institute (former NCA) ; the European Science Foundation (ESF, EU/QLRT-2001-01254), the European Community's Seventh Framework Program (FP7- HEALTH-F4-2007-2013, grant 01413: ENGAGE and grant 602768: ACTION); the European Research Council (ERC Starting 284167, ERC Consolidator 771057, ERC Advanced 230374), Rutgers University Cell and DNA Repository (NIMH U24 MH068457-

06), the National Institutes of Health (NIH, R01D0042157-01A1, R01MH58799-03, MH081802, DA018673, R01 DK092127-04, Grand Opportunity grants 1RC2 MH089951, and 1RC2 MH089995); the Avera Institute for Human Genetics, Sioux Falls, South Dakota (USA). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. Computing was supported by NWO through grant 2018/EW/00408559, BiG Grid, the Dutch e-Science Grid and SURFSARA.

OATS: The Older Australian Twins Study (OATS) is funded by NHMRC Program Grant ID1045325 and the NHMRC/Australian Research Council Strategic Award (ID401162). Twins Research Australia was supported by the NHRMC Enabling Grant 310667. We also thank the OATS participants and Research Team.

Osaka: Osaka study was supported by the Brain Mapping by Integrated Neurotechnologies for Disease Studies (Brain/MINDS: Grant Number JP18dm0207006), Brain/MINDS & beyond studies (Grant Number JP19dm0307002) and Health and Labor Sciences Research Grants for Comprehensive Research on Persons with Disabilities (Grant Number H26-seishin-ippan-012) from the Japan Agency for Medical Research and Development (AMED), Grants-in-Aid for Scientific Research (KAKENHI; Grant Number JP25293250 and JP16H05375). Some computations were performed at the Research Center for Computational Science, Okazaki, Japan.

PAFIP: The PAFIP study was supported by Instituto de Salud Carlos III, FIS 00/3095, 01/3129, PI020499, PI060507, PI10/00183, the SENY Fundació Research Grant CI 2005-0308007, and the Fundación Marqués de Valdecilla API07/011. Biological samples from our cohort were stored at the Valdecilla Biobank and genotyping services were conducted at the Spanish “Centro Nacional de Genotipado” (CEGEN-ISCIII).

MCIC/COBRE: The study is funded by the National Institutes of Health studies R01EB006841, P20GM103472, and P30GM122734, Department of Energy DE-FG02-99ER62764

PING: Data collection and sharing for the Pediatric Imaging, Neurocognition and Genetics (PING) Study (National Institutes of Health Grant RC2DA029475) were funded by the National Institute on Drug Abuse and the Eunice Kennedy Shriver National Institute of Child Health & Human Development. A full list of PING investigators is at http://pingstudy.ucsd.edu/investigators.html.

QTIM: The QTIM study was supported by the National Institute of Child Health and Human Development (R01 HD050735), and the National Health and Medical Research Council (NHMRC 486682, 1009064),

Australia. Genotyping was supported by NHMRC (389875). Medland is supported in part by an NHMRC fellowship (APP1103623).

SHIP: SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide single-nucleotide polymorphism typing in SHIP and MRI scans in SHIP and SHIP-TREND have been supported by a joint grant from Siemens Healthineers, Erlangen, Germany and the Federal State of Mecklenburg-West Pomerania.

StrokeMRI: StrokeMRI was supported by the Norwegian ExtraFoundation for Health and Rehabilitation (2015/FO5146), the Research Council of Norway (249795, 262372), the South-Eastern Norway Regional Health Authority (2014097, 2015044, 2015073), and the Department of Psychology, University of Oslo.

Sydney MAS: The Sydney Memory and Aging Study (Sydney MAS) is funded by National and Health Medical Research Council (NHMRC) Program and Project Grants (ID350833, ID568969, ID109308). We also thank the Sydney MAS participants and the Research Team.

SYS: The SYS Study is supported by Canadian Institutes of Health Research.

TOP: Centre of Excellence: RCN #23273. RCN #. 226971. Part of this work was performed on the TSD (Tjeneste for Sensitive Data) facilities, owned by the University of Oslo, operated and developed by the TSD service group at the University of Oslo, IT-Department (USIT). (tsd-drift@usit.uio.no). The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7-PEOPLE-2013-COFUND) under grant agreement n° 609020 - Scientia Fellows; the Research Council of Norway (RCN) # 276082 - A lifespan perspective on mental illness: toward precision medicine using multimodal brain imaging and genetics. . IES is supported by South-Eastern Norway Regional Health Authority (#2020060), European Union's Horizon2020 Research and Innovation Programme (CoMorMent project; grant #847776) and the KG Jebsen Foundation (SKGJ-MED-021).

UCLA_UMCU: The UCLA_UMCU cohort comprises of 6 studies which were supported by National Alliance for Research in Schizophrenia and Affective Disorders (NARSAD) (20244 to Prof. Hillegers), The Netherlands Organisation for Health Research and Development (ZonMw) (908-02-123 to Prof. Hulshoff Pol), and Netherlands Organisation for Scientific Research (NWO 9120818 and NWO- VIDI 917-46-370 to Prof. Hulshoff Pol). The GROUP study was funded through the Geestkracht programme of the Dutch Health Research Council (ZonMw, grant number 10-000-1001), and matching funds from

participating pharmaceutical companies (Lundbeck, AstraZeneca, Eli Lilly, Janssen Cilag) and universities and mental health care organizations (Amsterdam: Academic Psychiatric Centre of the Academic Medical Center and the mental health institutions: GGZ inGeest, Arkin, Dijk en Duin, GGZ Rivierduinen, Erasmus Medical Centre, GGZ Noord Holland Noord. Groningen: University Medical Center Groningen and the mental health institutions: Lentis, GGZ Friesland, GGZ Drenthe, Dimence, Mediant, GGNet Warnsveld, Yulius Dordrecht, and Parnassia psycho-medical center, The Hague. Maastricht: Maastricht University Medical Centre and the mental health institutions: GGzE, GGZ Breburg, GGZ Oost-Brabant, Vincent van Gogh voor Geestelijke Gezondheid, Mondriaan, Virenze riagg, Zuyderland GGZ, MET ggz, Universitair Centrum Sint- Jozef Kortenberg, CAPRI University of Antwerp, PC Ziekeren Sint-Truiden, PZ Sancta Maria Sint-Truiden, GGZ Overpelt, OPZ Rekem. Utrecht: University Medical Center Utrecht and the mental health institutions Altrecht, GGZ Centraal, and Delta.)

UK Biobank: This work made use of data sharing from UK Biobank (under project code 27412).

Pierre Vanderhaeghen: This work was funded by the Vlaams Instituut voor Biotechnologie (VIB), the European Research Council (ERC Adv Grant GENDEVOCORTEX), the Belgian FWO and FRS/FNRS, the AXA Research Fund, Ikuo Suzuki was supported by a postdoctoral Fellowship of the FRS/FNRS.

Conflicts of Interest

Dr Brodaty is an advisory board member for Nutricia Australia. He has received research funding from the EU "Joint Programme Neurodegenerative Disorders (JPND) and the National Health and Medical Research Council, Australia.

Dr Grabe has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm and Janssen Cilag. He has received research funding from the German Research Foundation (DFG), the German Ministry of Education and Research (BMBF), the DAMP Foundation, Fresenius Medical Care, the EU "Joint Programme Neurodegenerative Disorders" (JPND) and the European Social Fund (ESF).

Dr Ching, Thompson and Jahanshad received partial grant support from Biogen, Inc. (Boston), for research unrelated to the topic of this manuscript.

825 Dr Crespo-Facorro has received honoraria from Janssen Cilag, Otsuka and Lundbeck for
826 educational and advisory/consultant activities unrelated with the present research.
827 Dr Dale is a Founder of and holds equity in CorTechs Labs, Inc, and serves on its
828 Scientific Advisory Board. He is also a member of the Scientific Advisory Board of
829 Human Longevity, Inc., and receives funding through a research agreement with General
830 Electric Healthcare (GEHC). The terms of these arrangements have been reviewed and
831 approved by the University of California, San Diego in accordance with its conflict of
832 interest policies.
833 Mr Walters, Drs Gústafsson, Ulfarsson, Hreinn Stefánsson and Kári Stefánsson are
834 employees of deCODE genetics (Amgen).
835 Dr Hibar is an employee of Genentech, Inc. Dr Andreassen has received speakers
836 honorarium from Lundbeck, and is a consultant to HealthLytix.
837 Dr Astri J. Lundervold, has received speakers honorarium from Shire.
838 Drs Owen, Hall and van den Bree report grants from Takeda Pharmaceuticals outside of
839 the submitted work.
840 Dr Stein has received research grants and/or consultancy honoraria from Lundbeck and
841 Sun.
842 Dr. Jacob Nielsen and Dr. Fejgin are employed by H. Lundbeck A/S.
843 Dr Linden receives editorial fees from Elsevier and book royalties from Springer Nature
844 and Oxford University Press.
845 Dr Fladby is on the Novo Nordisk advisory board.
846 All other authors declare no competing financial interests.
847

848 **References**

849 1. Toga AW, Thompson PM. Genetics of brain structure and intelligence. *Annual*
850 *review of neuroscience* 2005; **28:** 1-23.

2. Hibar DP, Adams HH, Jahanshad N, Chauhan G, Stein JL, Hofer E *et al.* Novel genetic loci associated with hippocampal volume. *Nature communications* 2017; **8:** 13624.

3. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N *et al.* Common genetic variants influence human subcortical brain structures. *Nature* 2015; **520**(7546)**:** 224-229.

4. Meda SA, Pryweller JR, Thornton-Wells TA. Regional brain differences in cortical thickness, surface area and subcortical volume in individuals with Williams syndrome. *PloS one* 2012; **7**(2)**:** e31913.

5. Fan CC, Brown TT, Bartsch H, Kuperman JM, Hagler DJ, Jr., Schork A *et al.* Williams syndrome-specific neuroanatomical profile and its associations with behavioral features. *NeuroImage Clinical* 2017; **15:** 343-347.

6. Lin A, Ching CRK, Vajdi A, Sun D, Jonas RK, Jalbrzikowski M *et al.* Mapping 22q11.2 Gene Dosage Effects on Brain Morphometry. *J Neurosci* 2017; **37**(26)**:** 6183-6199.

7. Sun D, Ching CRK, Lin A, Forsyth JK, Kushan L, Vajdi A *et al.* Large-scale mapping of cortical alterations in 22q11.2 deletion syndrome: Convergence with idiopathic psychosis and effects of deletion size. *Mol Psychiatry* 2018.

8. Stefansson H, Meyer-Lindenberg A, Steinberg S, Magnusdottir B, Morgen K, Arnarsdottir S *et al.* CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature* 2014; **505**(7483)**:** 361-366.

9. Silva AI, Ulfarsson MO, Stefansson H, Gustafsson O, Walters GB, Linden DEJ *et al.* Reciprocal White Matter Changes Associated With Copy Number Variation at 15q11.2 BP1-BP2: A Diffusion Tensor Imaging Study. *Biol Psychiatry* 2018.

10. Ulfarsson MO, Walters GB, Gustafsson O, Steinberg S, Silva A, Doyle OM *et al.* 15q11.2 CNV affects cognitive, structural and functional correlates of dyslexia and dyscalculia. *Translational psychiatry* 2017; **7**(4)**:** e1109.

11. van der Meer D, Sonderby IE, Kaufmann T, Walters GB, Abdellaoui A, Ames D *et al.* Association of Copy Number Variation of the 15q11.2 BP1-BP2 Region With Cortical and Subcortical Morphology and Cognition. *JAMA psychiatry* 2019**:** 1-11.

12. Maillard AM, Ruef A, Pizzagalli F, Migliavacca E, Hippolyte L, Adaszewski S *et al.* The 16p11.2 locus modulates brain structures common to autism, schizophrenia and obesity. *Mol Psychiatry* 2015; **20**(1)**:** 140-147.

13. Qureshi AY, Mueller S, Snyder AZ, Mukherjee P, Berman JI, Roberts TP *et al.* Opposing brain differences in 16p11.2 deletion and duplication carriers. *J Neurosci* 2014; **34**(34)**:** 11199-11211.

14. Martin-Brevet S, Rodriguez-Herreros B, Nielsen JA, Moreau C, Modenato C, Maillard AM *et al.* Quantifying the effects of 16p11.2 copy number variants on brain structure: A multi-site 'genetic-first' study. *Biological Psychiatry* 2018.

15. Sonderby IE, Gustafsson O, Doan NT, Hibar DP, Martin-Brevet S, Abdellaoui A *et al.* Dose response of the 16p11.2 distal copy number variant on intracranial volume and basal ganglia. *Mol Psychiatry* 2018; **25**(3)**:** 584-602.

16. Kendall KM, Rees E, Escott-Price V, Einon M, Thomas R, Hewitt J *et al.* Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. *Biol Psychiatry* 2016.

17. Huguet G, Schramm C, Douard E, Jiang L, Labbe A, Tihy F *et al.* Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA psychiatry* 2018; **75**(5)**:** 447-457.

18. Hippolyte L, Maillard AM, Rodriguez-Herreros B, Pain A, Martin-Brevet S, Ferrari C *et al.* The Number of Genomic Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. *Biol Psychiatry* 2016; **80**(2)**:** 129-139.

19. Bernier R, Steinman KJ, Reilly B, Wallace AS, Sherr EH, Pojman N *et al.* Clinical phenotype of the recurrent 1q21.1 copy-number variant. *Genetics in medicine : official journal of the American College of Medical Genetics* 2016; **18**(4)**:** 341-349.

20. Brunetti-Pierri N, Berg JS, Scaglia F, Belmont J, Bacino CA, Sahoo T *et al.* Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* 2008; **40**(12)**:** 1466-1471.

21. Rosenfeld JA, Traylor RN, Schaefer GB, McPherson EW, Ballif BC, Klopocki E *et al.* Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes. *European journal of human genetics : EJHG* 2012; **20**(7)**:** 754-761.

22. Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A *et al.* Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* 2018; **173**(6)**:** 1370-1384.e1316.

23. Fiddes IT, Lodewijk GA, Mooring M, Bosworth CM, Ewing AD, Mantalas GL *et al.* Human-Specific NOTCH2NL Genes Affect Notch Signaling and Cortical Neurogenesis. *Cell* 2018; **173**(6)**:** 1356-1369.e1322.

24. PGC PCaSWGot. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet* 2017; **49**(1)**:** 27-35.

25. Haldeman-Englert CR, Jewett T. 1q21.1 Recurrent Microdeletion. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K *et al.* (eds). *GeneReviews((R))*. University of Washington, Seattle University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.: Seattle (WA), 1993.

26. Mefford HC, Sharp AJ, Baker C, Itsara A, Jiang Z, Buysse K *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *The New England journal of medicine* 2008; **359**(16)**:** 1685-1699.

27. Chawner S, Owen MJ, Holmans P, Raymond FL, Skuse D, Hall J *et al.* Genotype-phenotype associations in children with copy number variants associated with high neuropsychiatric risk in the UK (IMAGINE-ID): a case-control cohort study. *The lancet Psychiatry* 2019; **6**(6)**:** 493-505.

28. Gourari I, Schubert R, Prasad A. 1q21.1 Duplication syndrome and epilepsy: Case report and review. *Neurology Genetics* 2018; **4**(1)**:** e219.

29. Gudmundsson OO, Walters GB, Ingason A, Johansson S, Zayats T, Athanasiu L *et al.* Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. *Translational psychiatry* 2019; **9**(1)**:** 258.

30. Green EK, Rees E, Walters JT, Smith KG, Forty L, Grozeva D *et al.* Copy number variation in bipolar disorder. *Mol Psychiatry* 2016; **21**(1)**:** 89-93.

31. Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S *et al.* Association of Rare Copy Number Variants With Risk of Depression. *JAMA psychiatry* 2019.

32. Kendall KM, Bracher-Smith M, Fitzpatrick H, Lynham A, Rees E, Escott-Price V *et al.* Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *The British journal of psychiatry : the journal of mental science* 2019; **214**(5)**:** 297-304.

33. Owen D, Bracher-Smith M, Kendall KM, Rees E, Einon M, Escott-Price V *et al.* Effects of pathogenic CNVs on physical traits in participants of the UK Biobank. *BMC genomics* 2018; **19**(1)**:** 867.

34. Mace A, Tuke MA, Deelen P, Kristiansson K, Mattsson H, Noukas M *et al.* CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. *Nature communications* 2017; **8**(1)**:** 744.

35. Crawford K, Bracher-Smith M, Owen D, Kendall KM, Rees E, Pardinas AF *et al.* Medical consequences of pathogenic CNVs in adults: analysis of the UK Biobank. *Journal of medical genetics* 2018.

36. Dolcetti A, Silversides CK, Marshall CR, Lionel AC, Stavropoulos DJ, Scherer SW *et al.* 1q21.1 Microduplication expression in adults. *Genetics in medicine : official journal of the American College of Medical Genetics* 2013; **15**(4)**:** 282-289.

37. Verhagen JM, de Leeuw N, Papatsonis DN, Grijseels EW, de Krijger RR, Wessels MW. Phenotypic Variability Associated with a Large Recurrent 1q21.1 Microduplication in a Three-Generation Family. *Molecular syndromology* 2015; **6**(2)**:** 71-76.

38. Soemedi R, Topf A, Wilson IJ, Darlay R, Rahman T, Glen E *et al.* Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. *Hum Mol Genet* 2012; **21**(7)**:** 1513-1520.

39. Buse M, Cuttaia HC, Palazzo D, Mazara MV, Lauricella SA, Malacarne M *et al.* Expanding the phenotype of reciprocal 1q21.1 deletions and duplications: a case series. *Italian journal of pediatrics* 2017; **43**(1)**:** 61.

40. Rosenfeld JA, Coe BP, Eichler EE, Cuckle H, Shaffer LG. Estimates of penetrance for recurrent pathogenic copy-number variations. *Genetics in medicine : official journal of the American College of Medical Genetics* 2013; **15**(6)**:** 478-481.

41. Nielsen J, Fejgin K, Sotty F, Nielsen V, Mork A, Christoffersen CT *et al.* A mouse model of the schizophrenia-associated 1q21.1 microdeletion syndrome exhibits altered mesolimbic dopamine transmission. *Translational psychiatry* 2017; **7**(11)**:** 1261.

42. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA *et al.* Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. *Nat Genet* 2006; **38**(9)**:** 1038-1042.

43. O'Bleness M, Searles VB, Dickens CM, Astling D, Albracht D, Mak AC *et al.* Finished sequence and assembly of the DUF1220-rich 1q21 region using a haploid human genome. *BMC genomics* 2014; **15:** 387.

44. Dumas LJ, O'Bleness MS, Davis JM, Dickens CM, Anderson N, Keeney JG *et al.* DUF1220-domain copy number implicated in human brain-size pathology and evolution. *Am J Hum Genet* 2012; **91**(3)**:** 444-454.

45. Florio M, Heide M, Pinson A, Brandl H, Albert M, Winkler S *et al.* Evolution and cell-type specificity of human-specific genes preferentially expressed in progenitors of fetal neocortex. *eLife* 2018; **7**.

46. Fiddes IT, Pollen AA, Davis JM, Sikela JM. Paired involvement of human-specific Olduvai domains and NOTCH2NL genes in human brain evolution. *Human genetics* 2019; **138**(7)**:** 715-721.

47. Harvard C, Strong E, Mercier E, Colnaghi R, Alcantara D, Chow E *et al.* Understanding the impact of 1q21.1 copy number variant. *Orphanet journal of rare diseases* 2011; **6:** 54.

48. Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J *et al.* Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* 2016; **19**(11)**:** 1523-1536.

49. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SF *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome research* 2007; **17**(11)**:** 1665-1674.

50. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag: New Yrok, 2009.

51. Rakic P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends in neurosciences* 1995; **18**(9)**:** 383-388.

52. Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP *et al.* The genetic architecture of the human cerebral cortex. *Science* 2020; **367**(6484).

53. Di Donato N, Chiari S, Mirzaa GM, Aldinger K, Parrini E, Olds C *et al.* Lissencephaly: Expanded imaging and clinical classification. *American journal of medical genetics Part A* 2017; **173**(6)**:** 1473-1488.

54. Libby J, Marghoub A, Johnson D, Khonsari RH, Fagan MJ, Moazen M. Modelling human skull growth: a validated computational model. *Journal of the Royal Society, Interface* 2017; **14**(130).

55. Warland A, Kendall KM, Rees E, Kirov G, Caseras X. Schizophrenia-associated genomic copy number variants and subcortical brain volumes in the UK Biobank. *Mol Psychiatry* 2019.

56. Adams HH, Hibar DP, Chouraki V, Stein JL, Nyquist PA, Renteria ME *et al.* Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nat Neurosci* 2016; **19**(12)**:** 1569-1582.

57. Hofer E, Roshchupkin GV, Adams HHH, Knol MJ, Lin H, Li S *et al.* Genetic correlations and genome-wide associations of cortical structure in general population samples of 22,824 adults. *Nature communications* 2020; **11**(1)**:** 4796.

58. Steinberg KM, Schneider VA, Graves-Lindsay TA, Fulton RS, Agarwala R, Huddleston J *et al.* Single haplotype assembly of the human genome from a hydatidiform mole. *Genome research* 2014; **24**(12)**:** 2066-2076.

59. Reinwald JR, Sartorius A, Weber-Fahr W, Sack M, Becker R, Didriksen M *et al.* Separable neural mechanisms for the pleiotropic association of copy number variants with neuropsychiatric traits. *Translational psychiatry* 2020; **10**(1)**:** 93.

60. Hill J, Inder T, Neil J, Dierker D, Harwell J, Van Essen D. Similar patterns of cortical expansion during human development and evolution. *Proceedings of the National Academy of Sciences* 2010; **107**(29)**:** 13135.

61. Fjell AM, Westlye LT, Amlien I, Tamnes CK, Grydeland H, Engvig A *et al.* High-expanding cortical regions in human development and evolution are related to higher intellectual abilities. *Cerebral cortex (New York, NY : 1991)* 2015; **25**(1)**:** 26-34.

62. Deary IJ, Penke L, Johnson W. The neuroscience of human intelligence differences. *Nature reviews Neuroscience* 2010; **11**(3)**:** 201-211.

63. Lodewijk GA, Fernandes DP, Vretzakis I, Savage JE, Jacobs FMJ. Evolution of human brain-size associated NOTCH2NL genes proceeds towards reduced protein levels. *Molecular biology and evolution* 2020.

64. Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LSJ *et al.* Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *The lancet Psychiatry* 2017; **4**(4)**:** 310-319.

65. van Rooij D, Anagnostou E, Arango C, Auzias G, Behrmann M, Busatto GF *et al.* Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group. *Am J Psychiatry* 2018; **175**(4)**:** 359-369.

66. van Erp TG, Hibar DP, Rasmussen JM, Glahn DC, Pearlson GD, Andreassen OA *et al.* Subcortical brain volume abnormalities in 2028 individuals with schizophrenia and 2540 healthy controls via the ENIGMA consortium. *Mol Psychiatry* 2016; **21**(4)**:** 547-553.

67. Hibar DP, Westlye LT, van Erp TG, Rasmussen J, Leonardo CD, Faskowitz J *et al.* Subcortical volumetric abnormalities in bipolar disorder. *Mol Psychiatry* 2016; **21**(12)**:** 1710-1716.

68. Schmaal L, Veltman DJ, van Erp TG, Samann PG, Frodl T, Jahanshad N *et al.* Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol Psychiatry* 2016; **21**(6)**:** 806-812.

69. Whelan CD, Altmann A, Botia JA, Jahanshad N, Hibar DP, Absil J *et al.* Structural brain abnormalities in the common epilepsies assessed in a

worldwide ENIGMA study. *Brain : a journal of neurology* 2018; **141**(2)**:** 391-408.

Tables

Table 1: Demographic data.

	ENIGMA-CNV				deCODE			
	del	**nc**	**dup**	**P**	**del**	**nc**	**dup**	**P**
n	28	37088	22		2	1150	5	
Sex = male (%)	15 (54%)	17912 (48%)	9 (41%)		1 (50%)	511 (44%)	2 (33%)	
Age (mean (sd))	41.7 (19.0)	61.1 (12.8)	55.4 (12.7)	<0.001	53.5 (2.1)	44.8 (12.4)	46.4 (16.5)	
Children (age<18)	4 (14%)	665 (1.8%)		<0.001	0	0	0	
Known diagnosis (%)	11 (39.3%)	2424 (6.5%)	7 (32%)	<0.001		238 (21%)	2 (40%)	
DiseaseType (%)								
ADHD		1 (~0%)				181 (16%)	2 (40%)	
Autism						2 (0.2%)		
Bipolar disorder						7 (0.6%)		
Clinically recruited (no diagnosis)	6 (21.4%)		4 (18%)					
Dyslexia	1 (3.6 %)							
F-ICD-10-diagnosis (UK biobank)		858 (2.3%)	1 (4%)					
G-ICD10-diagnosis (UK biobank)	1 (3.0%)	1439 (3.8%)	1 (4%)					
MDD		1 (~0%)						
Multiple diagnoses*	2 (7.2%)		1 (4.5%)					
Persistent depressive disorder		1 (~0%)						
SCZ	1 (3.6)	124 (0.3)				48 (4.2%)		
Scannersites	11	15	8		2	2	1	
Datasets	9	13	7		1	1	1	

P (p-value) is based on a Chi-squared test for categorical values and ANOVA for continuous values.ADHD = attention deficit disorder, clinically recruited = recruited in clinical NDD study but without a diagnosis, MDD = major depressive disorder, SCZ = schizophrenia, del = deletion carrier, nc = non-carriers, dup = duplication carrier, NS=non-significant. * [1st deletion carrier: Agoraphobia, Avoidant personality disorder (AvPD), Obsessive Compulsive Disorder (OCD), Dependent personality disorder (DPD), other substance related disorder, conduct disorder. 2nd deletion carrier: Specific phobia, social phobia, MDD, AvPD, Schizotypal personality disorder (STPD). Duplication carrier: Social phobia, OCD, MDD, AvPD.]

Table 2: 1q21.1 CNV deletion and duplication carriers show deficits in specific cognitive functions.

		n			del vs nc			dup vs nc		
Test	**Suggested domain**	**del**	**nc**	**dup**	**Cohen's D (SE)**	**P**		**Cohen's D (SE)**	**P**	
Pairs Matching	Working memory	119	468 709	186	**-0.36 (0.09)**	**7.3E-05**	**	0.03 (0.01)	0.7	
Reaction Time	Simple processing speed	115	464 648	181	-0.12 (0.06)	0.18		**-0.23 (0.07)**	**2.1E-03**	*
Reasoning and problem solving	Fluid Intelligence	29	154 490	71	-0.48 (0.19)	9.2E-03		**-0.33 (0.12)**	**5.3E-03**	*
Digit Span	Numeric memory	12	47 569	27	-0.27 (0.14)	0.36		0.14 (0.07)	0.47	
Symbol Digit Substitution	Complex processing speed	24	111 900	28	**-0.78 (0.2)**	**1.4E-04**	**	0.04 (0.02)	0.83	
Trail Making A	Visual attention	23	98 495	27	-0.29 (0.15)	0.16		-0.14 (0.07)	0.45	
Trail Making B	Visual attention	23	98 494	27	**-0.87 (0.21)**	**3.1E-05**	***	-0.19 (0.1)	0.33	

Note: n=sample size; del=Deletion carriers; dup=duplication carriers; nc=non-carriers; SE=Standard error, P=P-value. Multiple comparison-corrected significant findings (p<.007) are indicated in bold and with *<0.007, **<0.0007, ***<0.00007.

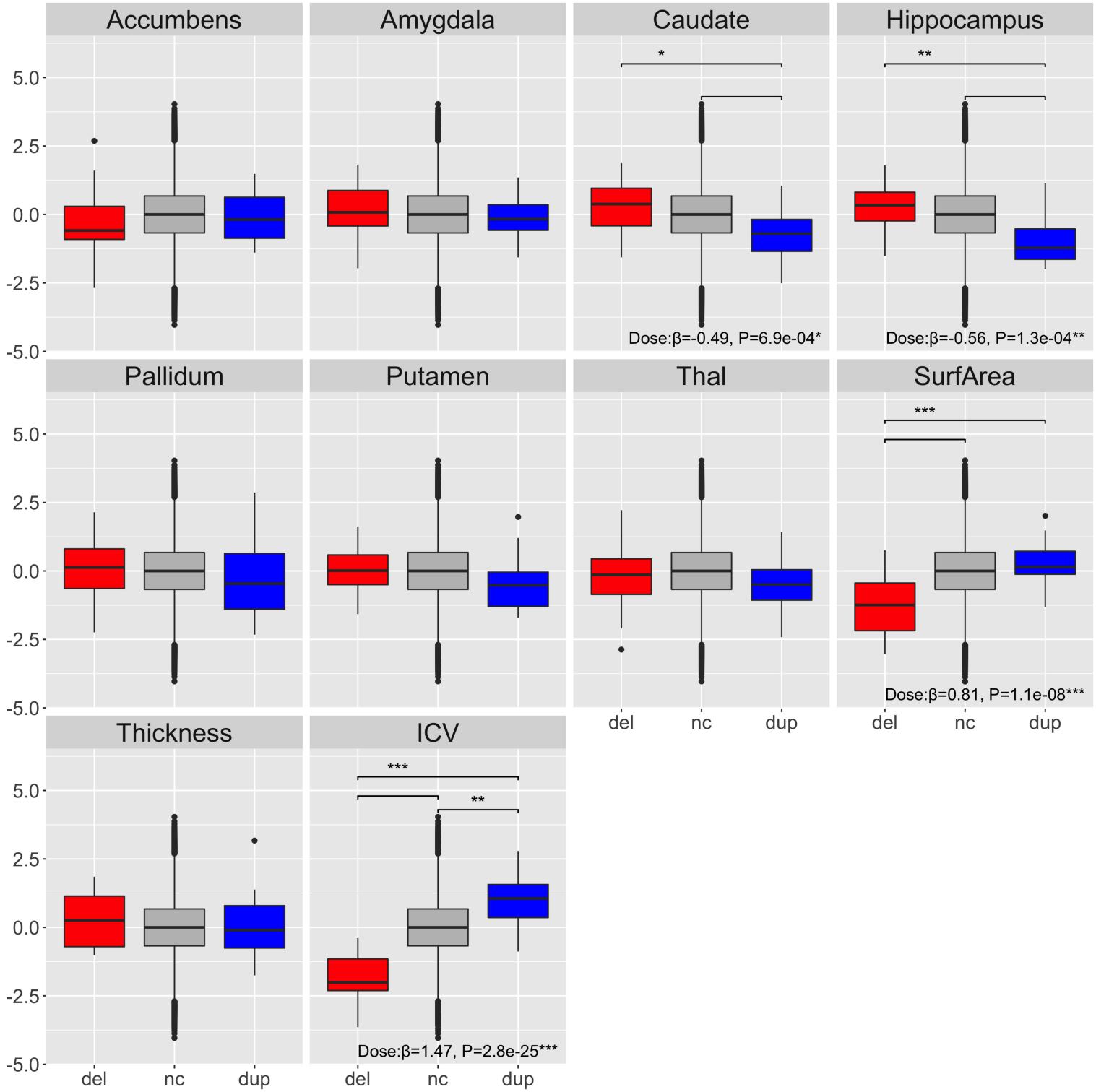
Table 3: Mediation analysis of brain structures over the association between 1q21.1 distal CNV carrier status and performance in the cognitive tasks in the UK Biobank. Path B is the effect of the brain structure on cognition overall including all 1q21.1 deletion and duplication carriers (4-13 CNV carriers in each group) and non-carriers (n= 10,501-30,924; for exact numbers, see Supplementary Table 13). Each calculation included 5000 simulations.

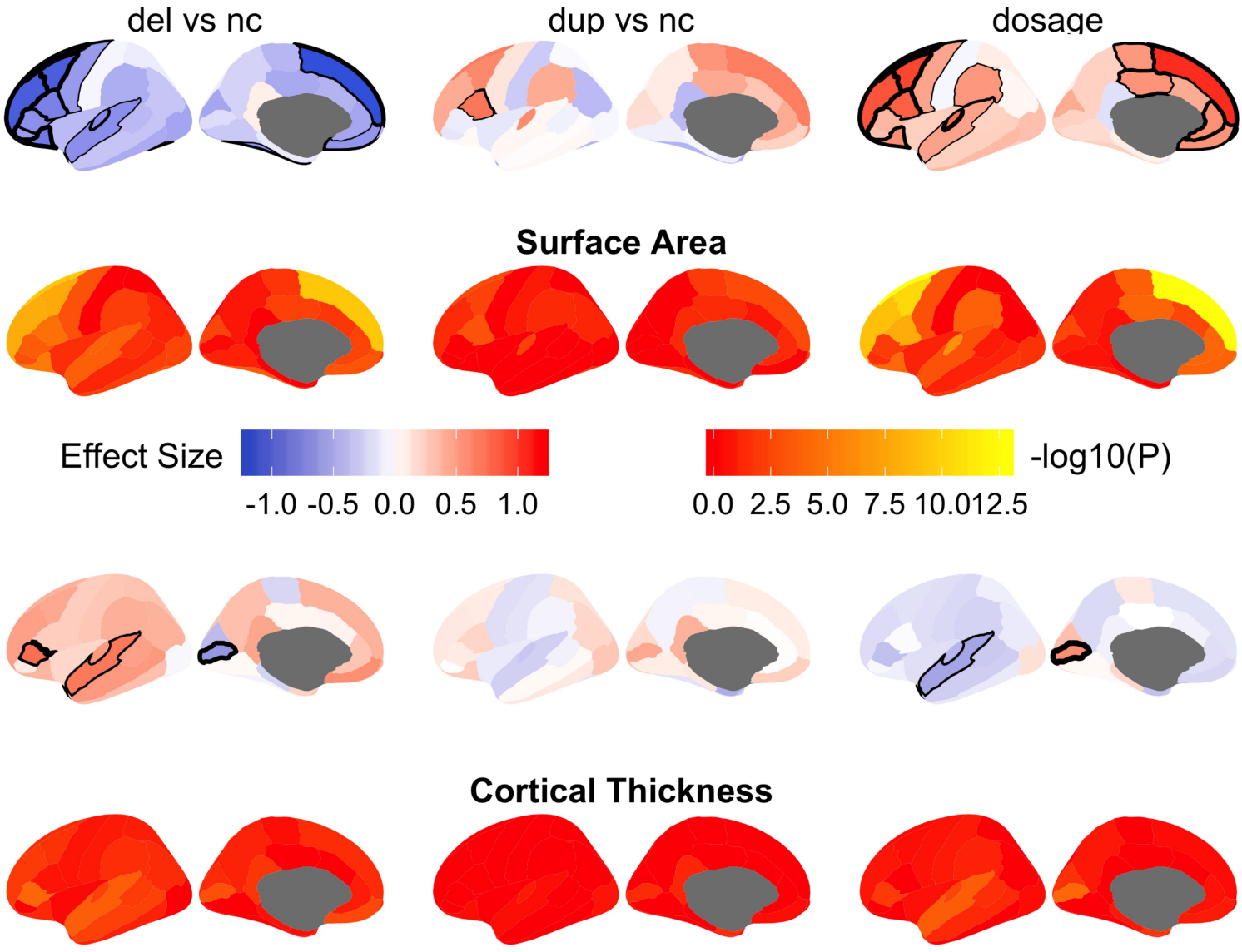
	Path B - effect of brain structure on cognition		DELETION		DUPLICATION	
	Estimate (SE)	**P**	**Prop. mediated**	**P**	**Prop. mediated**	**P**
Pairs matching						
Caudate	0.0023 (0.0053)	0.66			3.5E-03	0.85
Hippocampus	0.005 (0.0052)	0.34			9.8E-03	0.68
SurfArea	**0.031 (0.0055)**	**1.9E-08**	-0.07	0.65	-4.4E-03	0.9
ICV	**0.027 (0.0054)**	**4.3E-07**	-0.12	0.64	-0.07	0.51
Reaction Time						
Caudate	-0.0016 (0.0054)	0.77			-2.3E-03	0.67
Hippocampus	0.01 (0.0053)	0.053			*0.01*	*0.04*
SurfArea	-0.0095 (0.0056)	0.091	0.02	0.13	7.3E-04	0.78
ICV	0.029 (0.0055)	**2.4E-07**	-0.1	0.07	**-0.03**	**2.4E-03**
Reasoning and problem solving						
Caudate	-0.0059 (0.0091)	0.51			5.7E-03	0.55
Hippocampus	0.0031 (0.0089)	0.73			-9.6E-05	0.95
SurfArea	**0.052 (0.0094)**	**2.6E-08**	0.06	0.250	-7.4E-04	0.97
ICV	**0.15 (0.0092)**	**3.7E-59**	0.25	0.24	*0.18*	*0.04*
Symbol digit substitution						
Caudate	0.0011 (0.0077)	0.88			-4.2E-03	0.83
Hippocampus	**0.04 (0.0075)**	**6.5E-08**			-0.01	0.82
SurfArea	**0.055 (0.0079)**	**3.8E-12**	**0.05**	**2.4E-03**	6.9E-04	0.99
ICV	**0.066 (0.0079)**	**3.6E-17**	**0.1**	**4.0E-04**	0.13	0.68
Trail Making A						
Caudate	**0.034 (0.0084)**	**5.7E-05**			4.4E-04	1
Hippocampus	**0.04 (0.0081)**	**1.0E-06**			3.0E-03	0.97
SurfArea	**0.046 (0.0086)**	**1.1E-07**	0.09	0.19	1.1E-03	0.98
ICV	**0.059 (0.0085)**	**6.1E-12**	0.21	0.20	-0.01	0.99
Trail Making B						
Caudate	0.021 (0.0083)	0.012			-0.01	0.79
Hippocampus	**0.04 (0.008)**	**6.9E-07**			-0.01	0.86
SurfArea	**0.082 (0.0085)**	**6.4E-22**	**0.07**	**8.0E-04**	8.9E-03	0.92
ICV	**0.11 (0.0084)**	**1.2E-36**	**0.17**	**1.2E-03**	0.16	0.73

Figure legends

Figure 1: **Cortical surface area and ICV show a positive dosage effect and caudate and hippocampus a negative dosage effect to copy number in the 1q21.1 distal region in our main sample (ENIGMA-CNV and UK biobank).** Boxplots of subcortical volumes, cortical surface area and mean cortical thickness and ICV are shown. Deletion carriers (del) in red, non-carriers (nc) in grey and duplication carriers (dup) in blue, respectively. The normalized brain values are presented. Boxplots represent the mean. Copy number dosage effect is noted at the bottom of each panel. Significant differences after correction between groups are noted as * = P < 0.0014, ** = P < 0.00014, *** = 0.000014. Centre line represents median, box limits are the upper and lower 25 % quartiles, whiskers the 1.5 interquartile range and the points are the outliers. All analyses were corrected for age, age squared, sex, scanner site and ICV (except for ICV).

Figure 2. Results from the t-tests and linear regression of 1q21.1 copy number variation on regional cortical surface area and cortical thickness. 1st and 3rd rows: Effect sizes (Cohen's d for the t-tests, beta coefficient for the dosage/linear regression). 2nd and 4th rows: Statistical significance in –log10 of the p-value. Significant areas in rows 1 and 3 are marked with black lines with increasing thickness for increasing significance (P<0.0014). The column names indicate the comparisons with del=deletion carriers, nc=non-carriers, dup=duplication carriers. All measures were corrected for age, age^2, sex, scanner site and ICV.


Accumbens
Amygdala
Caudate
Hippocampus
Pallidum
Putamen
Thal
SurfArea
Thickness
ICV
5.0
2.5
0.0
-2.5
-5.0
*
**
***
Dose:β=-0.49, P=6.9e-04*
Dose:β=-0.56, P=1.3e-04**
Dose:β=0.81, P=1.1e-08***
Dose:β=1.47, P=2.8e-25***
del
nc
dup


del vs nc
dup vs nc
dosage
Surface Area
Effect Size
-1.0 -0.5 0.0 0.5 1.0
0.0 2.5 5.0 7.5 10.012.5
-log10(P)
Cortical Thickness

Supplementary Figures and Notes for: "1q21.1 distal copy number variants are associated with cerebral and cognitive alterations in humans"

Overview, Supplementary Figures (pages 3-12):

Overview, Supplementary Notes (pages 13-20):

Supplementary Note 1: Extended information on datasets.

Supplementary Note 2: Details on CNV calling and QC.

Supplementary Note 3: Extended information on UK biobank CNV calls.

Supplementary Note 4: Extended info on image acquisition and processing.

Supplementary Note 5: Description of additional sensitivity and robustness analyses.

Supplementary Note 6: Details on cognitive task data processing

Supplementary Note 7: Details on human fetal transcriptional data

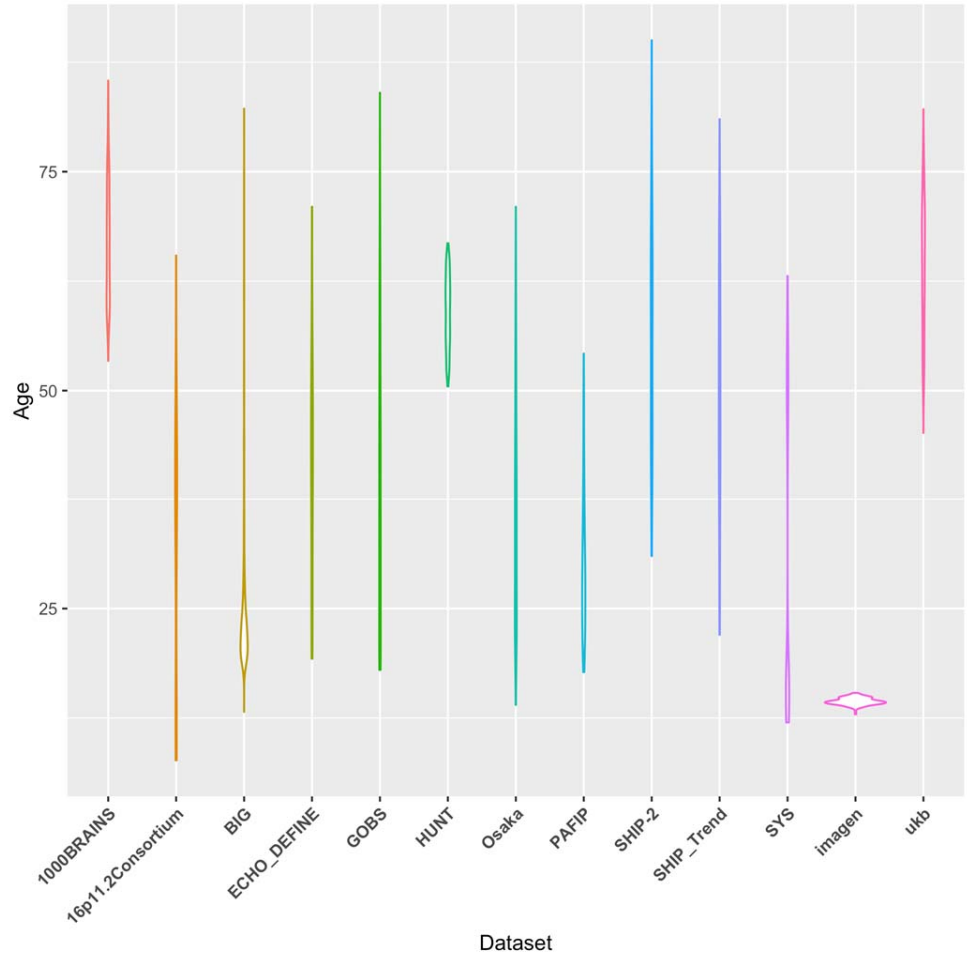
Supplementary Note 8: Df(h1q21)+/- mouse characterization

Supplementary Note 9: Results on the 1q21.1 distal deletion mouse

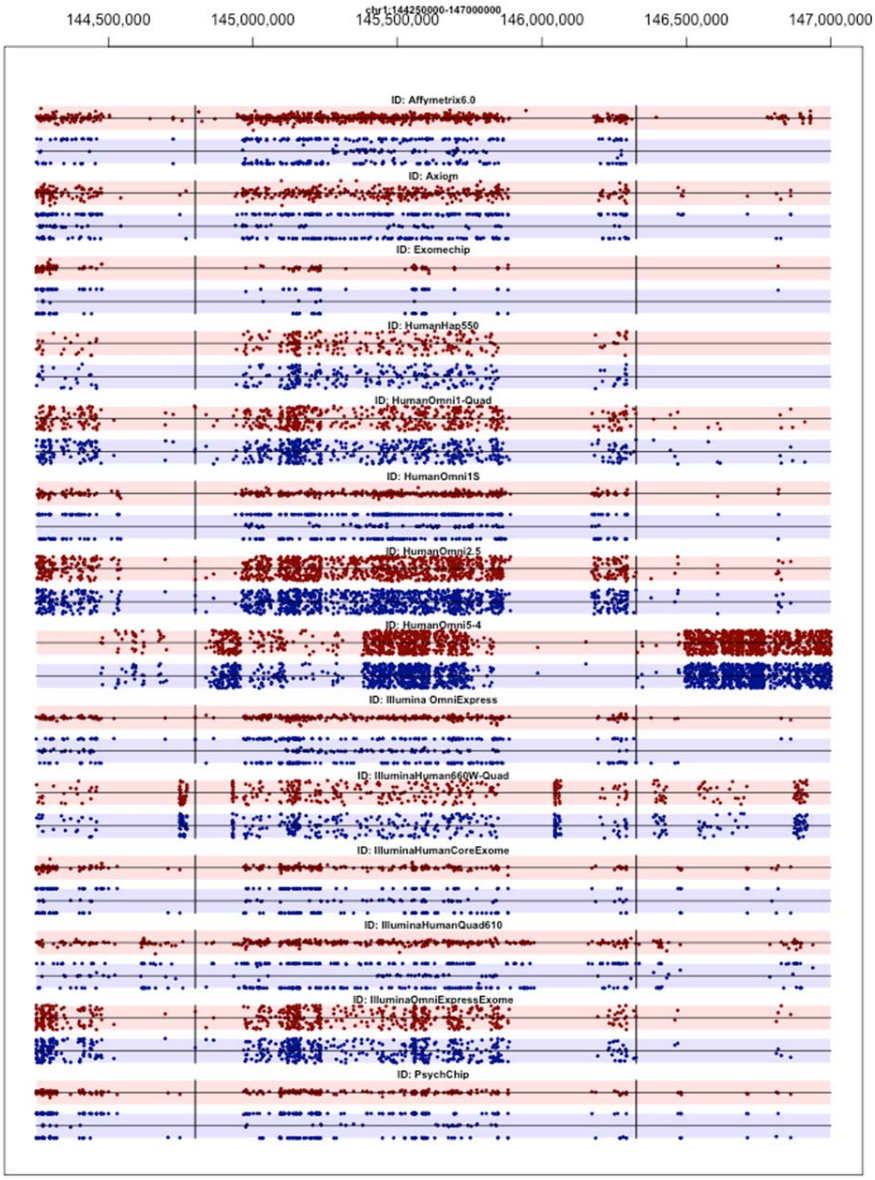
Overview, Supplementary tables, legends (pages 21-26):

(NOTE – this is ONLY legends - please refer to separately submitted excel sheet for the entire tables)

SUPPLEMENTARY FIGURES:

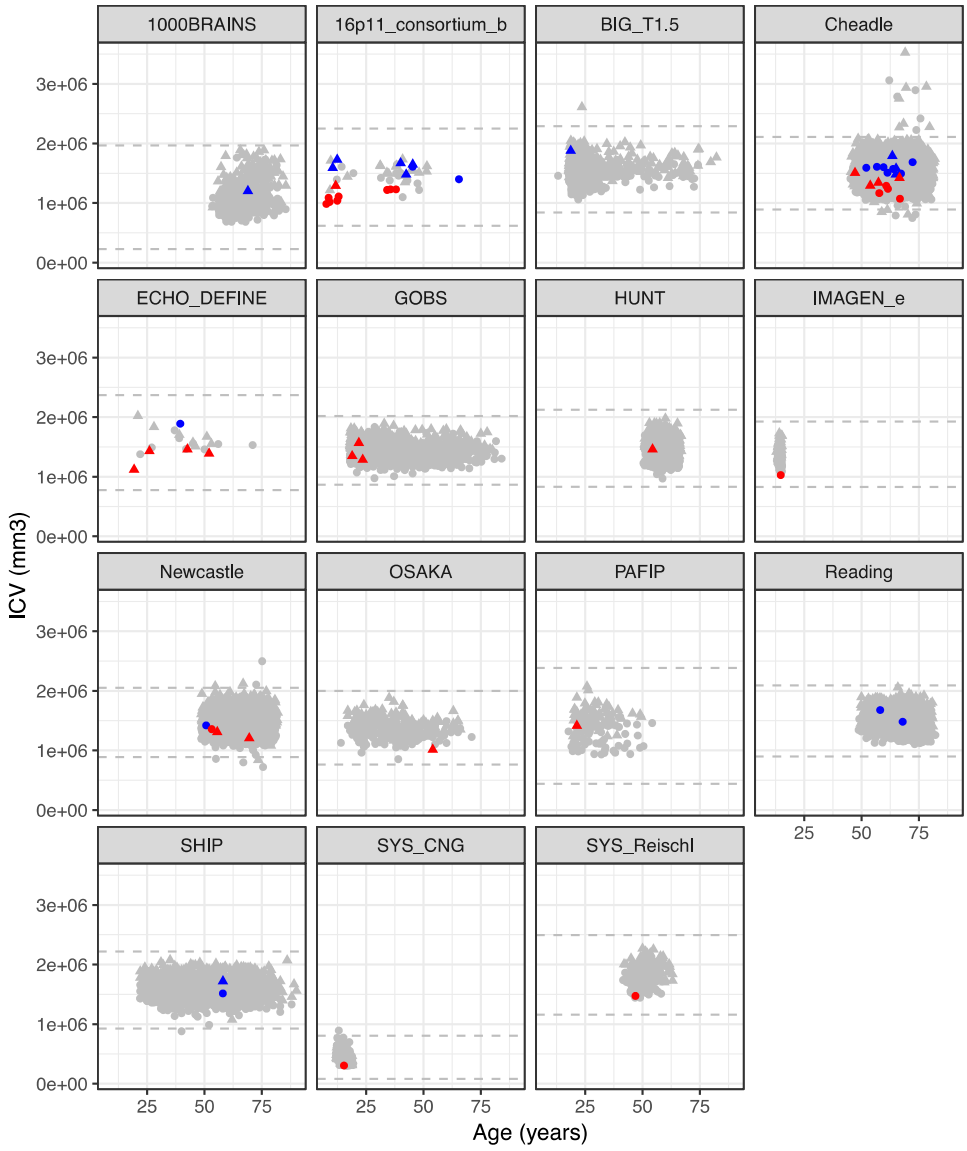


Supplementary Figure 1: Age distribution per cohort contributing data to the current study, with age in years on the y-axis and cohort name on the x-axis.

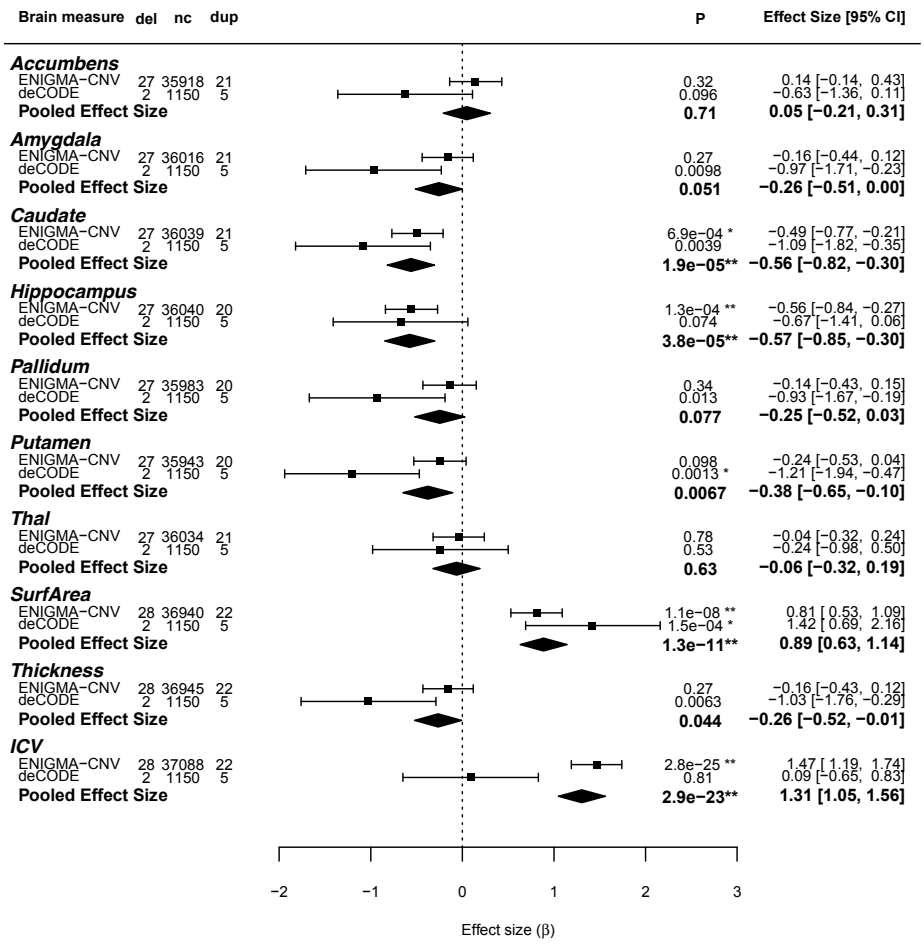


Supplementary Figure 2: Coverage of the 1q21.1 distal region by genotyping platforms in ENIGMA-CNV. Log R ratio is shown in red, B-allele frequency in blue. The vertical black lines delimit the boundaries of the 1q21.1 distal region. HumanHap550, HumanOmniQuad1-Quad, HumanOmni2.5, HumanOmni5-4,

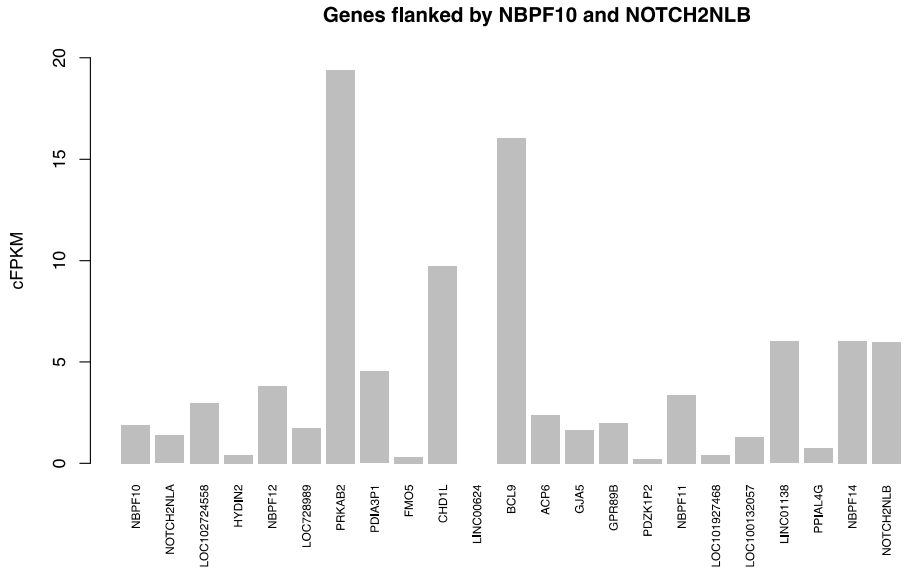
IlluminaHuman660-Quad, IlluminaOmniExpressExome are mock data. The rest is based on real data.



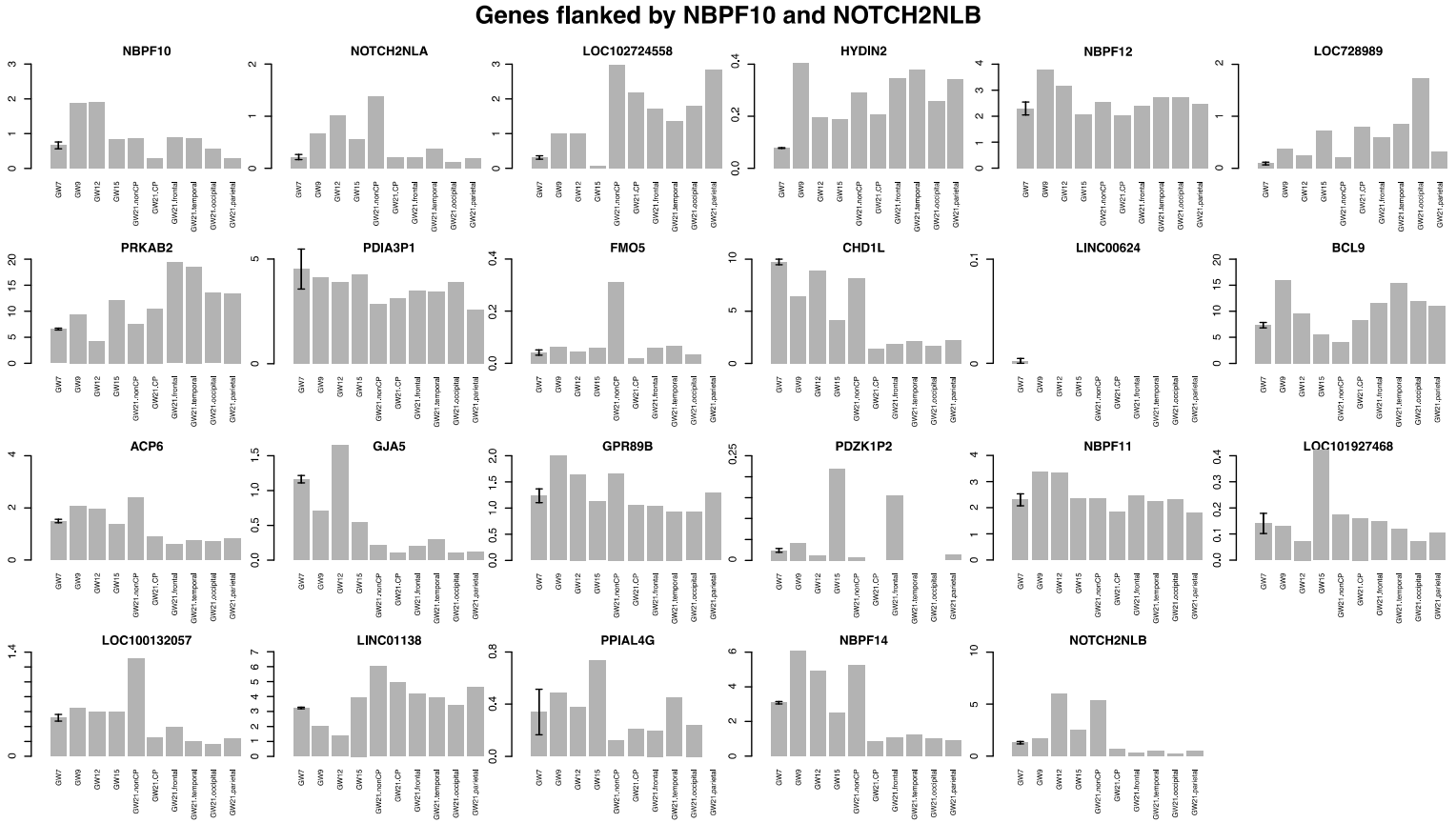
Supplementary Figure 3: Bivariate plot of age (years) versus uncorrected ICV (mm3). Deletion carriers in red, non-carriers in grey and duplication carriers in blue, respectively. Circles = females, triangles = males.



Supplementary Figure 4. Forest plots on the dosage effect of copy number on subcortical volumes, surface area, thickness and ICV. The effect size (β of the linear regression) at each site for each measure is shown by the position on the x-axis. Standard error is shown by the horizontal line. A summary polygon shows the results when fitting a random-effects model to the two datasets: the main and the Icelandic deCODE samples. del, nc and dup denote the number of individuals in each analysis. This number changes on the basis of quality control for each structure. * = P < 0.0014, ** = P < 0.00014. Effect size and confidence intervals are to the right.

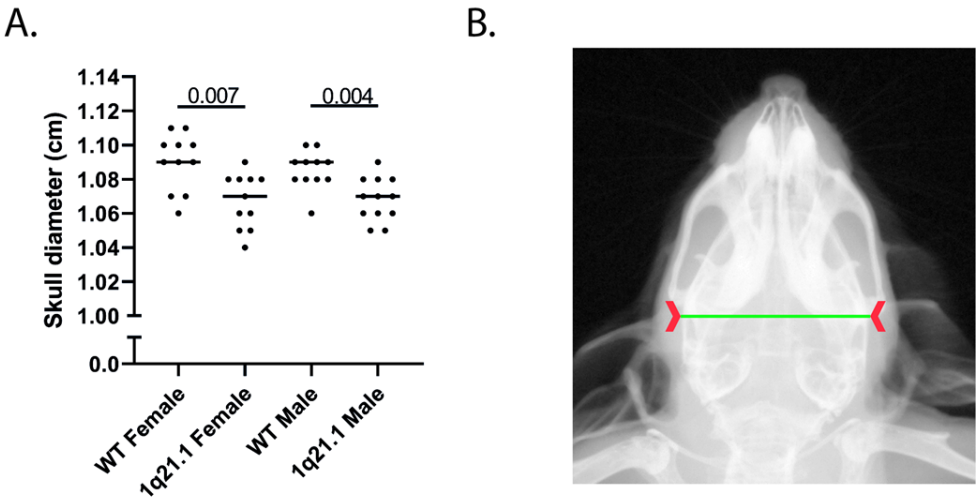


Supplementary Figure 5: Expression peak of the genes encoded in the 1q21.1 interval during human fetal corticogenesis. The highest expression value for each gene in the developmental stages from GW7 to GW21 is indicated. The genes are ordered according to their chromosomal positions. cFPKM = corrected Fragment Per Kilobase and Million reads.

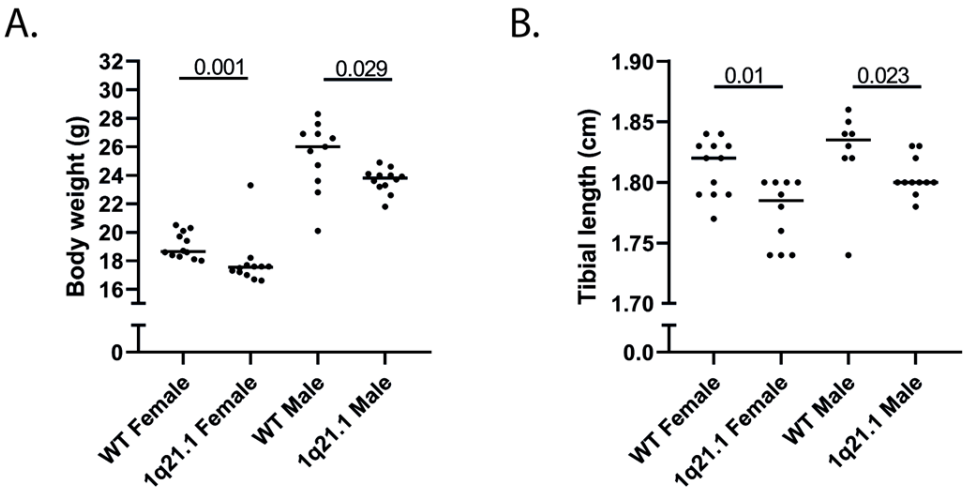


Supplementary Figure 6: RNA-seq profile of genes in the 1q21.1 interval during human corticogenesis. The expression value for each gene in corrected Fragment Per Kilobase and Million reads (cFPKM) in the developmental stages from gestation week 7 (GW7) to 21 (GW21) is

indicated. The GW21 stage samples include dissection into frontal, temporal, occipital and parietal lobes. The parietal area was further microdissected into the cortical plate (CP) and underlying domains of the cortical wall (non-CP, containing mostly outer-subventricular zone (oSVZ) and ventricular zone germinal zones.

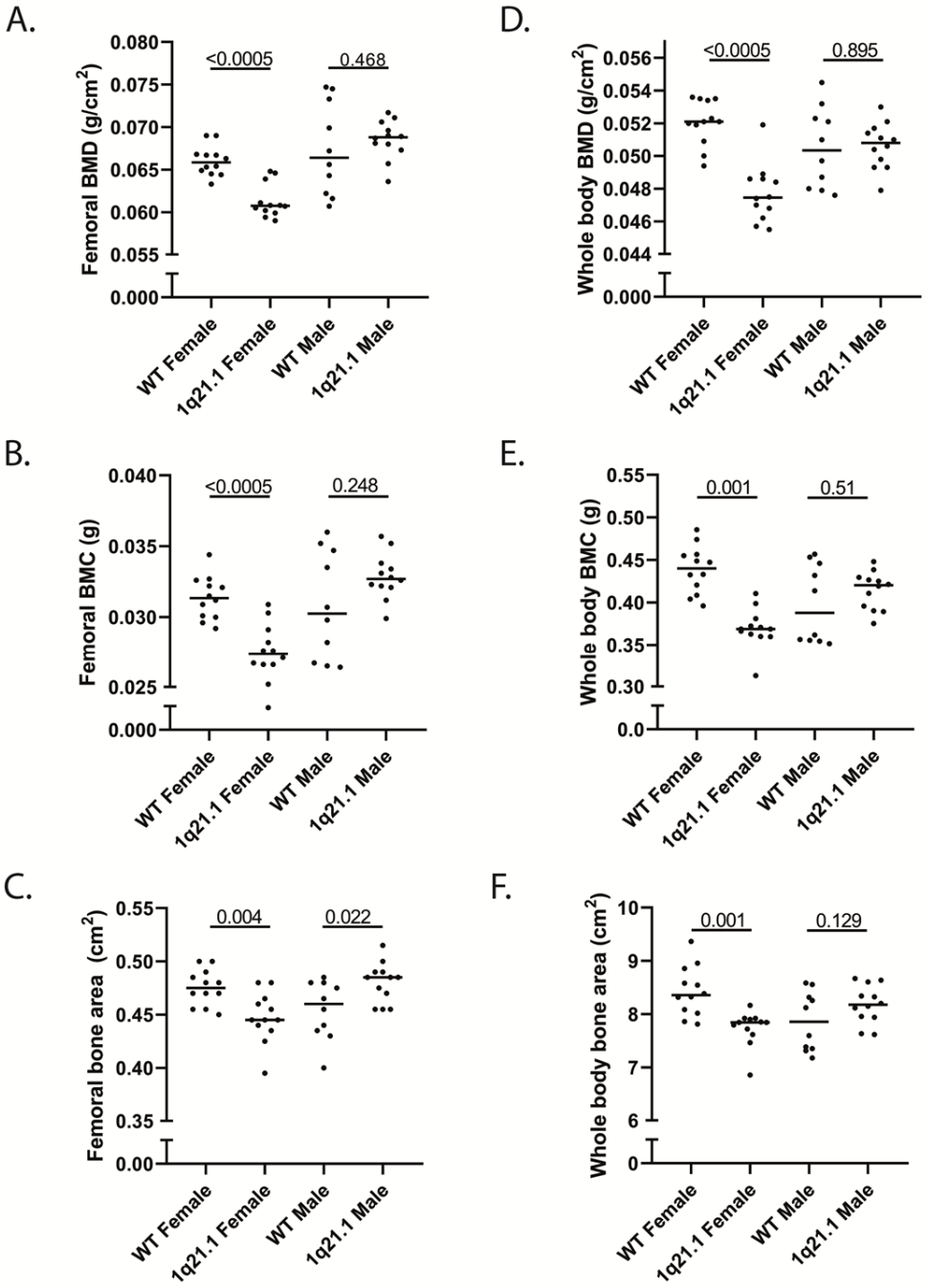


Supplementary Figure 7: Skull diameter in 1q21.1 deletion knockout mice in comparison to wildtype (WT) littermates. A. Median skull diameter (n=10-12 per group). The horizontal lines demark P-values as group-wise comparisons (non-parametric Mann-Whitney U test) between the genotype groups. B. X-ray showing the mouse skull with the green line indicating how the skull diameter was determined.



Supplementary Figure 8: Body weight and bone size of 1q21.1 deletion mice in

comparison to wildtype (WT) littermates. A. Median body weight. B. Tibial (lower leg) length measured on X-rays. P-values show group-wise comparisons (non-parametric Mann-Whitney U test) between the genotype groups (n=10-12 in each group).



Supplementary Figure 9: Bone mass measurements in 1q21.1 mice and wildtype (WT) litter mates. Median bone mineral density (BMD) in femur (upper leg) (A) and whole body (D). Median bone mineral content (BMC) in femur (B) and whole body (E). Bone area in femur (C) and whole body (F). P-values show group-wise comparisons (non-parametric Mann-Whitney U test) between the genotype groups.

SUPPLEMENTARY NOTES

Supplementary Note 1: Extended information on datasets and test for differences in demographics.

Diagnosis-information in ENIGMA-CNV was based on information from the different datasets. In the UK biobank, diagnosis was extracted as Datafield 41202: Diagnoses - main ICD10 and 41204 Diagnoses - secondary ICD10. If either of these contained an F (mental) or G (neurological) disorder, these were coded as affected and the ICD10 disorder was noted.

For the core ENIGMA-CNV dataset, family information was based on pi-hat estimated for pairs of individuals, and only one relative (if more than two) from pairs with pi-hat >0.2 was kept. CNV carriers were selectedly kept over non-carriers. For ECHO_DEFINE and the 16p11.2 European Consortium, relatedness was based on information from the clinican. For the UK biobank, relatives were extracted from Datafield 22011: Genetic relatedness pairing and Datafield 22012: Genetic relatedness factor. One of each pair with a kinship coefficient above 0.053 (that is more related than 1st cousins) was removed.

Tests for differences between groups for demographic data applied a test included in the R package tableone v0.7.3 – chi square test with continuity correction for categorical values and ANOVA for continuous variables.

Supplementary Note 2: Details on CNV calling and QC. All PFB-files were based on Human Genome Build NCBI36/hg18 except for UK biobank, ECHO-DEFINE and parts of 16p11.2 European Consortium that used NCBI37/hg19. PFB-and GC-files were selected based on publicly available data from the PennCNV homepage or self-generated: in the case of cohorts primarily consisting of Asian and African

individuals, a PFB-file was generated through PennCNV compile_pfb.pl using all genotyping arrays from the cohort. The PFB-file used is noted in Supplementary Table 10.

The following quality control metrics were used: Adjacent CNVs separated by a gap less than 20% of the combined length of the two CNVs were merged until no more gaps of <20% existed, and CNVs based on less than 15 SNPs were excluded. Only samples with standard deviation (SD) of normalized intensity (LRR) <0.35, B allele frequency (BAF) drifting value <0.01 and wave factor value between -0.05 and 0.05 were included.

The 1q21.1 distal region was well-covered by all arrays (Figure S8). CNVs overlapping the region of interest (1q21.1 distal and 1q21.1 distal and proximal) were identified with the R package iPsychCNV SelectSamplesFromROI with parameters OverlapMin = 0.4 and OverlapMax = 5, visualized with iPsychCNV StackPlot and manually inspected. None of the 1q21.1 distal carriers carried additional genomic imbalances (Supplementary Table 1) except for three duplications that extended into the 1q21.1 proximal region (Supplementary Table 2), a known susceptibility factor for thrombocytopenia-absent radius (TAR) syndrome[1]. In the statistical analysis, individuals with a minimum overlap of 0.4 to regions with known pathogenic CNVs (Table S10) were excluded regardless of copy number status as were individuals from scanner sites without 1q21.1 distal CNV carriers.

Carriers in the 16p11.2 European consortium cohort were identified based on report from the cytogeneticist who did the genetic test in the clinic and was thus based on either CGH array or FISH (Fluorescent In Situ Hybridization) - the identification method for each individual carrier is noted in Supplementary Table 2. Non-carriers in

the 16p11.2 European consortium cohort were either selected from the general population (excluding individuals with a neurodevelopmental or psychiatric diagnosis) or familiar controls who tested negative for the 1q21.1 distal and proximal CNV or familial controls from a 16p11.2 proximal and distal CNV study - five of the latter had a neuropsychiatric diagnosis. Carriers in the ECHO_DEFINE were identified based on the report from the cytogenetist after genetic test in the clinic with Psych Chip.

Supplementary Note 3: Extended information on UK biobank CNV calls.

Anonymised genotyped data was downloaded as l2r & baf-files from UK biobank showcase for chromosomes 1-22, X, Y, M & XY. In addition, snp-files were downloaded. They were stored and processed on a secure Unix server.

For the initial steps, the l2r- and baf-files were split into separate files for each individual containing both l2r and baf-values in 20 batches, each containing 25,000 individuals per batch [the last batch contained 13,377). Subsequently, SNP-names were added to the files. CNVs were called in subbatches of 1000 individuals per batch using PennCNV [57] and self-generated PFB- and GCC-model files (NCBI37/hg19) and affygw6.hmm. Subsequent filtering and visualization was done as for the main dataset above except that the LRR_SD cut-off was set at 0.50 given that we observed reliable CNV calls within these ranges. We did not filter based on number of CNVs or genotype call rate. These are quite relaxed filtering criteria but since all 1q21.1 CNVs were visualized and inspected and thus filtered for false positives, we did not apply more stringent parameters. 59 individuals were excluded from the entire UK biobank using these criteria.

Supplementary Note 4: Extended info on image acquisition and processing. Each site contributed volumes for the left and right hemispheres of the accumbens, caudate, putamen, pallidum, amygdala, hippocampus and thalamus in addition to right and left 34 regional cortical surface areas and 34 cortical average thicknesses, total surface area and total mean cortical thickness as well as estimated intracranial volume (ICV). The total volume, surface area or mean thickness of each structure was calculated by adding the left and right together. We excluded each individual measure if it deviated more than +/- 4SD from the mean for each individual scanner site.

Supplementary Note 5: Description of additional sensitivity and robustness analyses.

We re-analysed the dataset in the following way: (a) MATCHED analysis: Matching each CNV carrier with one non-carrier. The R package Matchit v2.4 was used to match each CNV carrier with one non-carrier based on sex, age, scanner site and ICV. (b) NON-AFFECTED only analysis: Keeping only non-affected individuals (i.e. individuals without a known diagnosis of a brain disorder), (c) NON-AFFECTED ADULTS analysis: Keeping only non-affected adults (age>=18) (d) ADULTS analysis: Only including adults with age>=18 (e) CHILDREN analysis: Only keeping children with age<18, or (f) ENIGMA-CNV ONLY: Keeping only ENIGMA-CNV derived individuals in analysis or (g) UK biobank ONLY: Keeping only UKB-derived participants or (h) POPULATION STRUCTURE analysis: Controlling for population structure by including 4 genetic principal components as covariates calculated based on standardized multidimensional scaling analyses of genome-wide genotype data conducted at each site (i) NO ICV model analysis: Excluding ICV as covariate or j)

(f) INCLUDING RELATIVES analysis: Including all relatives (first- or second-degree relatives) that was removed in the primary analysis.

Supplementary Note 6: Details on cognitive task data processing

The *Pairs Matching* task (field 399), tested episodic memory, with six pairs of cards being shown for three seconds to participants, before being turned over, after which the participants were asked to identify the matching pairs. We used the total number of errors made. The *Reaction Time* task (field 20023), tested simple processing speed through twelve rounds of a game where participants had to click a button as quickly as possible when shown two matching cards. We used the mean reaction time. *Fluid Intelligence* (field 20016), tested reasoning and problem solving through thirteen verbal and numerical reasoning questions, which had to be answered within two minutes. We used the total number of correct answers. The *Digit Span* task (field 4282) tested numeric working memory by presenting progressively longer numbers to participants and asking them to recall these once the number had disappeared. We used the maximum number of digits correctly recalled. The Symbol Digit Substitution task (field 20195) tested complex processing speed through the matching of numbers to a set of symbols. We used the number of correct substitutions. The Trail Making A and B tasks (fields 20156 and 20157) tested visual attention by asking participants to connect scattered circles according to numbers (trail A) and to alternating numbers and letters (trail B). We used the time taken to complete these tests for our analyses. All data was recoded so that higher scores indicate higher performance.

Supplementary Note 7: Details on human fetal transcriptional data

Human fetal tissue collection and preparation was done as described previously[19] - human fetuses were obtained following medical pregnancy termination. Fetuses aged

7 gestational weeks (GW) (2 males), 9 GW (1 male, 1 undetermined), 12 GW (1 female, 1 undetermined), 15 GW (1 male), and 21 GW (1 male) were used for the RNA sequencing and *in situ* hybridization of cortical tissue. All cases were examined with standard feto-pathological procedures and none displayed clinical or neuropathological evidence of brain malformation. The brain was removed within 6 hours of expulsion and RNA extracted and cDNA prepared[2]. The 350-700bp size cDNA fraction was sequenced from both ends using Hiseq 2500 Rapid mode v3 (Illumina). Transcriptome analysis was performed as previously described[2]. Expression values are calculated as the unit Fragment per kilobase and million reads (FPKM) and those for human-specific duplicated genes (*NBPF10, NOTCH2NLA, HYDIN2, NBPF12, LOC728989, NBPF11, NBPF14, and NOTCH2NLB*) are corrected on the basis of the computer simulation performed in the previous study (cFPKM; corrected FPKM)[2]. The study was approved by three relevant Ethics Committees (Erasme Hospital, Université Libre de Bruxelles, and Belgian National Fund for Scientific Research FRS/FNRS) on research involving human subjects. Written informed consent was given by the parents in each case.

Supplementary Note 8: Df(h1q21)+/- mouse characterization

16-week-old male and female heterozygous 1q21.1 deletion knockout mice[3] and wildtype mice were sacrificed for bone analysis (n=10-12 in each genetic group). Body weight was recorded. Femur and tibia were collected and stored in ethanol at 4 C and in saline at -20 C until further analysis. Animals were genotyped from Taconic (Ejby, Denmark) and genotyping was repeated on tail samples collected after sacrifice. Whole body DXA scans were obtained using a Piximus densitometer (GE Lunar, Madison, WI, USA). Whole-body and femoral bone mineral density (BMD),

bone mineral content and bone area were analysed. *X*-ray scans of the skull, upper limbs and lower limbs were obtained using a Faxitron MX-20 small animal x-ray system (Faxitron, Tucson, AZ, USA). Skull diameter, femur length and width, and humerus and tibia length were measured using the ruler function in a dicom viewer program. Measurements were done by staff blinded to genotype groups. 1q21 and wildtype mice were compared using non-parametric Mann-Whitney U test. To take variation in bone turnover between sexes into account, males and female mice (10-12 mice in analytical group) were analyzed separately. Differences were considered significant at $P<0.05$. All animals were included in the analyses.

Supplementary Note 9: Results on the the 1q21.1 deletion mouse

In a comparison between 1q21.1 deletion mice (Df(h1q21) +/- mouse[3])) and their wild-type littermates (n=10-12 mice per group), we found a significant decrease in skull diameter in the deletion mice (2% decrease, P=0.007 (females) and P=0.004 (males)) (Figure 4). Also, the deletion mice displayed lower weight and shorter tibial (lower leg) length (P=0.01 (females), P = 0.023 (males)) (Supplementary Figure 4). Finally, bone mineral density (BMD), bone mineral content (BMC) and bone area were lower in female 1q21.1 deletion mice compared to wild-type littermates ($P<0.0005$ (BMD and BMC) and P=0.004 (area)) whereas male deletion mice - unlike the deletion females - displayed an increased femoral bone (upper leg) area (P=0.022; Supplementary Figure 5).

References:

1. Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA *et al*. Compound inheritance of a low-frequency regulatory SNP and a rare null

mutation in exon-junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 2012; **44**(4)**:** 435-S432.

2. Suzuki IK, Gacquer D, Van Heurck R, Kumar D, Wojno M, Bilheu A *et al*. Human-Specific NOTCH2NL Genes Expand Cortical Neurogenesis through Delta/Notch Regulation. *Cell* 2018; **173**(6)**:** 1370-1384.e1316.

3. Nielsen J, Fejgin K, Sotty F, Nielsen V, Mork A, Christoffersen CT *et al*. A mouse model of the schizophrenia-associated 1q21.1 microdeletion syndrome exhibits altered mesolimbic dopamine transmission. *Translational psychiatry* 2017; **7**(11)**:** 1261.

SUPPLEMENTARY TABLES LEGENDS:

(NOTE – this is ONLY legends - please refer to separately submitted excel sheet for the entire tables)

Supplementary Table 1: Specification of all cohorts in ENIGMA CNV. Study design, participant demographics, and references to articles containing descriptions of individual inclusion and exclusion parameters for all datasets in ENIGMA-CNV collected up until Sep 30 2019. Data sets contributing data to the 1q21.1 distal analysis are marked with a star.

Supplementary Table 2: CNVs of Interest. Individuals with a minimum overlap of 0.4 to these CNVs were excluded from the analysis. Coordinates are Human Genome Build NCBI36/hg18 and GRCh37/hg19.

Supplementary Table 3: Chips and corresponding PFB-files used for PennCNV CNV calling.

Supplementary Table 4: Technical details concerning scanners and acquisition parameters utilized at the participating ENIGMA-CNV scanner sites

Supplementary Table 5: Sensitivity analyses – dosage effect of 1q21.1 distal copy number on subcortical volumes in the main sample. The effect size (β of the linear regression) is presented with 95 % confidence intervals. A linear regression based on the copy number state of the individuals (deletion=1, normal=2, duplication=3) was performed on normalized brain measures corrected for *plusICV*: age, age squared,

sex, scanner site and ICV (except for ICV) or *noICV*: age, age squared, sex, scanner site. Analysis was performed on : ALL – all individuals, ADULTS – adults (age ≥18), NON-AFFECTED - individuals without a known diagnosis of a brain disorder, NON-AFFECTED ADULTS – adult individuals without a known diagnosis of a brain disorders, MATCHED CONTROLS - matching each carrier with one non-carrier based on age, sex and scanner site or matching each carrier with one non-carrier based on age, sex, scanner site and ICV, POPULATION STRUCTURE – checking effect of population structure on individuals. Only individuals with accessible ancestry information were included in the analysis. ENIGMA-CNV ONLY – ENIGMA-CNV dataset exclusively. UKB ONLY – UK biobank dataset only, INCLUDING RELATIVES– including relatives with more than third degree relationships. Results were considered statistically significant if they were below a Bonferroni-corrected P-value of 0.0014. * = P < 0.0014, ** = P < 0.00014, ***=P<0.000014.

Supplementary Table 6: Sensitivity analyses - T-tests on subcortical volumes between different 1q21.1 distal copy number groups in the main sample. The effect size (Cohen's D) including 95 % confidence interval is presented. T-tests were performed on normalized values of brain measures *plusICV*: age, age squared, sex, scanner site and ICV (except for ICV) or *noICV*: age, age squared, sex, scanner site. Analysis was performed on: ALL – all individuals, ADULTS – adults (age ≥18), CHILDREN – children (age<18 years), CHILDREN – children (age<18 years), NON-AFFECTED - individuals without a known diagnosis of a brain disorder, NON-AFFECTED ADULTS – adult individuals without a known diagnosis of a brain disorders, MATCHED CONTROLS - matching each carrier with one non-carrier based on age, gender and scannersite or matching each carrier with one non-carrier

based on age, sex, scannersite and ICV, POPULATION STRUCTURE – checking effect of population structure on individuals. Only individuals with accessible ancestry information were included in the analysis, ENIGMA-CNV ONLY – ENIGMA-CNV dataset exclusively. UKB ONLY – UK biobank dataset only, INCLUDING RELATIVES– including relatives with more than third degree relationships. Results were considered statistically significant if they were below a Bonferroni-corrected P-value of 0.0014. * = P < 0.0014, ** = P < 0.00014, ***=P<0.000014.

Supplementary Table 7: Sensitivity analyses - 1q21.1 distal dosage effect on regional cortical surface area and mean cortical thickness. The effect size (β of the linear regression) is presented with 95 % confidence interval. A linear regression based on the copy number state of the individuals (deletion=1, normal=2, duplication=3) was performed on normalized brain measures corrected for *plusICV*: age, age squared, sex, scanner site and ICV (except for ICV) or *noICV*: age, age squared, sex, scanner site. Analysis was performed on all individuals with measures available. Analysis was performed on: ALL – all individuals, ADULTS – adults (age ≥18), NON-AFFECTED - individuals without a known diagnosis of a brain disorder, NON-AFFECTED ADULTS – adult individuals without a known diagnosis of a brain disorders, MATCHED CONTROLS - matching each carrier with one non-carrier based on age, gender and scanner site or matching each carrier with one non-carrier based on age, sex, scanner site and ICV, POPULATION STRUCTURE – checking effect of population structure on individuals. Only individuals with accessible ancestry information were included in the analysis NO RELATIVES– excluding relatives with more than third degree relationships. ENIGMA-CNV ONLY –

ENIGMA-CNV dataset exclusively. UKB ONLY – UK biobank dataset only, INCLUDING RELATIVES– including relatives with more than third degree relationships. Results were considered statistically significant if they were below a Bonferroni-corrected P-value of 0.0014. * = P < 0.0014, ** = P < 0.00014, ***=P<0.000014.

Supplementary Table 8: Sensitivity analyses: T-tests on regional cortical surface area and mean cortical thickness. The effect size (Cohen's D) including 95 % confidence interval is presented. T-tests were performed on normalized values of brain measures *plusICV*: age, age squared, sex, scanner site and ICV (except for ICV) or *noICV*: age, age squared, sex, scanner site. Analysis was performed on all individuals with measures available. Analysis was performed on: ALL – all individuals, ADULTS – adults (age ≥18), CHILDREN – children (age<18 years), CHILDREN – children (age<18 years), NON-AFFECTED - individuals without a known diagnosis of a brain disorder, NON-AFFECTED ADULTS – adult individuals without a known diagnosis of a brain disorders, MATCHED CONTROLS - matching each carrier with one non-carrier based on age, gender and scannersite or matching each carrier with one non-carrier based on age, sex, scannersite and ICV, POPULATION STRUCTURE – checking effect of population structure on individuals. Only individuals with accessible ancestry information were included in the analysis. ENIGMA-CNV ONLY – ENIGMA-CNV dataset exclusively. UKB ONLY – UK biobank dataset only, INCLUDING RELATIVES– including relatives with more than third degree relationships. Results were considered statistically significant if they were below a Bonferroni-corrected P-value of 0.0014. * = P < 0.0014, ** = P < 0.00014, ***=P<0.000014.

Supplementary Table 9: Demographic details of ENIGMA-CNV and UK biobank separately.

Supplementary Table 10: Extended information on 1q21.1 distal carriers. Established diagnosis (1 = yes, 0 = no), DiseaseType = known diagnosis or type of study. Chip = genotyping chip used for CNV calling. No of rels = number of known relatives in dataset. Relative = relative in dataset, relative removed = whether individual was removed from the analysis without relatives.

Supplementary Table 11: Meta-analysis of dosage effect of 1q21.1 distal copy number on subcortical volumes. The effect size (β of the linear regression) is presented. A linear regression based on the copy number state of the individuals (deletion (del) =1, non-carrier (nc) =2, duplication (dup)=3) was performed on normalized brain measures correcting for age^2, age, sex and scannersite (and ICV)) in the ENIGMA-CNV and UK biobank (main sample) and the independent Icelandic cohorts. A final effect size estimate of the combined sample was obtained using a fixed effects meta-analysis framework (metafor). Results were considered statistically significant if they were below a Bonferroni-corrected P-value of 0.0014. * = P < 0.0014, ** = P < 0.00014, ***=P < 0.000014, CI = confidence. Q = statistics for the test for heterogeneity, p(Q) = p-value for the test for heterogeneity, I2 =heterogeneity levels.

Supplementary Table 12: Meta-analysis of t-tests on subcortical volumes between different 1q21.1 distal copy number groups. The effect size (Cohen's D)

including 95 % confidence interval is presented. T-tests were performed on normalized values of brain measures correcting for age^2, age, sex and scanner site (and ICV) in the ENIGMA-CNV and UK biobank (main sample) and the independent Icelandic cohorts. A final effect size estimate of the combined sample was obtained using a fixed effects meta-analysis framework (metafor). Results were considered statistically significant if they were below a Bonferroni-corrected P-value of 0.0014. * = P < 0.0014, ** = P < 0.00014, ***=P < 0.000014, CI = confidence. Q = statistics for the test for heterogeneity, p(Q) = p-value for the test for heterogeneity, I2 =heterogeneity levels.

Supplementary Table 13: Available sample sizes per task, per carrier group, for the analyses linking the neuroimaging measures to the cognitive measures. In the analyses, we included all 1q21.1 CNV carriers and non-carriers in the UK Biobank with data on the seven cognitive tasks and brain structures.