

Title

10Kin1day: A bottom-up neuroimaging initiative

Authors

Martijn P van den Heuvel #1,2, Lianne H Scholtens 1,2, Hannelore K van der Burgh 3, Fredica Agosta 4, Clara Alloza 5, Tiffany M C Avancini 6, Simon Baron-Cohen 7, Silvia Basaia 4, Manon JNL Benders 8, Frauke Beyer 9, Linda Booij 10, Kees PJ Braun 11, Wiepke Cahn 2, Dara M Cannon 12, Sandra SM Chan 13, Eric YH Chen 14, Cambridge Child Development Project Consortium includes: Auyeung, Bonnie; Holt, Rosemary J; Lombardo, Michael V 7, Benedicto Crespo-Facorro 15, Eveline A Crone 16, Udo Dannlowski 17, Sonja MC de Zwarte 2, Ana M Diaz-Zuluaga 18, Bruno Dietsche 19, Stefan Du Plessis 20, Sarah Durston 2, Robin Emsley 20, Geraldo B Filho 6, Massimo Filippi 4, Thomas Frodl 21, Dariusz Gąsecki 22, Joanna Goc 23, Martin Gorges 24, Beata Graff 25, Dominik Grotegerd 17, Julie M Hall 26, Laurena Holleran 27, Helene J Hopman 28, Lutz Jäncke 29, Andreas Jansen 19, Krzysztof Jodzio 30, Vasiliy G Kaleda 31, Jan Kassubek 24, Shahrzad Kharabian Masouleh 9, Tilo Kircher 19, Martijn GJC Koevoets 2, Vladimir S Kostic 32, Axel Krug 19, Stephen M Lawrie 5, Irina S Lebedeva 33, Edwin HM Lee 14, Tristram A Lett 34, Simon JG Lewis 26, Franziskus Liem 35, Carlos Lopez-Jaramillo 18, Daniel S Margulies 36, Sebastian Markett 37, Paulo Marques 38, Ignacio Martínez-Zalacaín 39, Colm McDonald 12, Andrew McIntosh 5, Genevieve McPhilemy 12, Susanne L Meinert 17, José M Menchón 39, Susan Mérillat 35, Christian Montag 40, Pedro S

Moreira 38, Pedro Morgado 38, David Omar Mothersill 27, Hans-Peter Müller 24,
Leila Nabulsi 12, Pablo Najt 12, Krzysztof Narkiewicz 25, Patrycja Naumczyk 30,
Sebastiaan WF Neggers 41, Bob Oranje 42, Victor Ortiz-Garcia de la Foz 15,
Jiska S Peper 16, Julian A Pineda Z. 43, Paul E Rasser 44, Ronny Redlich 17,
Jonathan Repple 17, Martin Reuter 37, Pedro GP Rosa 6, Amber NV Ruigrok 7,
Agnieszka Sabisz 45, Ulrich Schall 44, Soraya Seedat 20, Mauricio H Serpa 6,
Devon Shook 42, Stavros Skouras 46, Carles Soriano-Mas 39, Nuno Sousa 38,
Edyta Szurowska 45, Alexander S Tomyshov 33, Diana Tordesillas-Gutierrez 47,
Leonardo Tozzi 48, Sofie L Valk 49, Leonard H van den Berg 3, Theo GM van
Erp 50, Neeltje EM van Haren 2, Judith JMC van Leeuwen 2, Arno Villringer 9,
Christiaan Vinkers 2, Christian Vollmar 23, Lea Waller 34, Henrik Walter 34,
Heather C Whalley 5, Marta Witkowska 30, Veronica A Witte 9, Marcus V Zanetti
6, Rui Zhang 9, Gary Donohoe 27, Veronica O'Keane 51, Siemon C de Lange
1,2

Corresponding author

Correspondence to Martijn P. van den Heuvel: Connectome Lab, Department of
Complex Trait Genetics, Center for Neurogenomics and Cognitive Research,
Vrije Universiteit Amsterdam, Amsterdam Neuroscience, De Boelelaan 1085,
1081 HV Amsterdam, The Netherlands; Department of Clinical Genetics,
Amsterdam UMC, Amsterdam Neuroscience, De Boelelaan 1117, 1081 HV
Amsterdam, The Netherlands; E-mail address: martijn.vanden.heuvel@vu.nl

Affiliations

1 Dutch Connectome Lab, Department of Neurosciences and Complex Traits Genetics, Center for Neurogenomics and Cognitive Research, VU Amsterdam and Department of Clinical Genetics, Amsterdam University Medical Center, Amsterdam, The Netherlands

2 Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands

3 Brain Center Rudolf Magnus, Department of Neurology, University Medical Center Utrecht, Utrecht, The Netherlands

4 Neuroimaging Research Unit, Institute of Experimental Neurology, Division of Neuroscience, San Raffaele Scientific Institute, Vita-Salute San Raffaele University, Milan, Italy

5 Division of Psychiatry, University of Edinburgh, Edinburgh, United Kingdom

6 Laboratory of Psychiatric Neuroimaging (LIM21), Departamento de Psiquiatria, Faculdade de Medicina, Universidade de São Paulo, Brazil. ; Núcleo de Apoio à Pesquisa em Neurociência Aplicada (NAPNA), Universidade de São Paulo, Brazil.

7 Autism Research Centre, Department of Psychiatry, University of Cambridge, United Kingdom

8 Brain Center Rudolf Magnus, Department of Neonatology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands

9 Department of Neurology, Max Planck Institute for Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany

10 Concordia University, Toronto, Canada

11 Brain Center Rudolf Magnus, Department of Child Neurology, Wilhelmina Children's Hospital, University Medical Center Utrecht, Utrecht, The Netherlands

12 Centre for Neuroimaging & Cognitive Genomics (NICOG), Clinical Neuroimaging Laboratory, NCBES Galway Neuroscience Centre, College of Medicine Nursing and Health Sciences, National University of Ireland Galway, H91 TK33 Galway, Ireland.

13 Department of Psychiatry, Chinese University of Hong Kong, Hong Kong SAR

14 Department of Psychiatry, University of Hong Kong, Hong Kong, China

15 Department of Medicine and Psychiatry. Psychiatry Unit. Hospital Universitario Marques de Valdecilla. IDIVAL,CIBERSAM, Santander,Spain

16 Brain and Development Research Center

17 Department of Psychiatry, University of Münster, Germany

18 Research Group in Psychiatry GIPSI, Department of Psychiatry, Faculty of Medicine, Universidad de Antioquia, Medellín, Colombia

19 Department of Psychiatry, University of Marburg, Germany

20 Department of Psychiatry, Stellenbosch University

21 Department of Psychiatry and Psychotherapy, University Hospital, Otto von Guericke University, Magdeburg, Germany

22 Department of Neurology of Adults, Medical University of Gdansk, Gdansk, Poland

23 Department of Neurology, Epilepsy Centre, University of Munich Hospital, Munich, Germany

24 Dept. of Neurology, University of Ulm, Germany

25 Department of Hypertension and Diabetology, Medical University of Gdansk, Gdansk, Poland

26 Parkinson's Disease Research Clinic, Brain and Mind Centre, University of Sydney, Sydney, Australia

27 Cognitive Genetics and Cognitive Therapy Group, Neuroimaging and Cognitive Genomics (NICOG) Centre & NCBES Galway Neuroscience Centre, School of Psychology and Discipline of Biochemistry, National University of Ireland Galway, Ireland

28 Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong SAR

29 Division of Neuropsychology, University of Zurich, Zurich, Switzerland

30 Institute of Psychology, University of Gdansk, Gdansk, Poland

31 Department of Endogenous Mental Disorders, Mental Health Research Center, Moscow, Russia

32 Clinic of Neurology, School of Medicine, University of Belgrade, Belgrade, Serbia

33 Laboratory of Neuroimaging and Multimodal Analysis, Mental Health Research Center, Moscow, Russia

34 Charité – Universitätsmedizin Berlin, Division of Mind and Brain Research, Department of Psychiatry and Psychotherapy, 10117 Berlin, Germany

35 University Research Priority Program "Dynamics of Healthy Aging", University of Zurich, Zurich, Switzerland

36 Max Planck Research Group for Neuroanatomy & Connectivity, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

37 Department of Psychology, University of Bonn, Germany

38 Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal

39 Department of Psychiatry, Bellvitge Biomedical Research Institute-IDIBELL and CIBERSAM, Barcelona, Spain

40 Department of Psychology, Ulm University, Germany

41 Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, UMC Utrecht, the Netherlands

42 Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, The Netherlands

43 Research Group, Instituto de Alta Tecnología Médica IATM, Medellín, Colombia

44 Priority Centre for Brain & Mental Health Research, The University of Newcastle, Australia

45 2nd Department of Radiology, Medical University of Gdańsk, Gdańsk, Poland

46 Barcelonabeta Brain Research Center, Barcelona, Spain

47 Neuroimaging Unit, Technological Facilities. Valdecilla Biomedical Research Institute IDIVAL, CIBERSAM, Santander, Cantabria, Spain

48 Department of Psychiatry and Psychotherapy, University Hospital, Otto von Guericke University, Magdeburg, Germany

49 Dept. of Social Neurosciences, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

50 Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, CA, USA

51 Division of Psychiatry, Trinity College, Dublin, Ireland

Abstract

We organized 10Kin1day, a pop-up scientific event with the goal to bring together neuroimaging groups from around the world to jointly analyze 10,000+ existing MRI connectivity datasets during a 3-day workshop. In this report, we describe the motivation and principles of 10Kin1day, together with a public release of 8,000+ MRI connectome maps of the human brain.

Main Text

Ongoing grand-scale projects like the European Human Brain Project (Amunts et al., 2016), the US Brain Initiative (Insel et al., 2013), the Human Connectome Project (Van Essen et al., 2013), the Chinese Brainnetome (Jiang, 2013) and exciting world-wide neuroimaging collaborations such as ENIGMA (Thompson et al., 2017) herald the new era of *big neuroscience*. In conjunction with these major undertakings, there is an emerging trend for bottom-up initiatives, starting with small-scale projects built upon existing collaborations and infrastructures. As described by Mainen and colleagues (Mainen et al., 2016), these initiatives are centralized around self-organized groups of researchers working on the same challenges and sharing interests and specialized expertise. These projects could scale and open up to a larger audience and other disciplines over time, eventually lining up and merging their findings with other programs to make the bigger picture.

10Kin1day

One type of event that fits well with this grass-roots collaboration philosophy are short gatherings of scientists around a single theme, bringing together expertise and tools to jointly analyze existing neuroscience data. We organized 10Kin1day, an MRI connectome event, with the goal to bring together an international group of researchers in the field of neuroimaging and consistently analyze MRI connectivity data of the human cerebrum. We organized the event around five founding principles:

- use existing neuroimaging data, available from many research groups around the world; we focused on diffusion MRI data and aimed to bring together 10,000+ datasets
- analyze data from varying cohorts and imaging protocols, using a single, straightforward analysis strategy to encourage across-group collaborations and multisite studies
- perform all processing during a short workshop, with only basic expertise of analysis needed
- provide education on how to analyze resulting connectome data, so participants can continue to work on their projects after the event
- each participant analyzes their own data and is free to decide what to do with their analyzed results

The 10K workshop

Over 50 participants from 40 different neuroimaging groups gathered in The Netherlands for a 3-day event. Participants brought and worked on their own datasets, varying from MRI data on healthy human brain organization, cross-sectional and longitudinal brain development, aging, cognitive psychology, as well as MRI data of a wide range of neurological and psychiatric brain disorders (including among others: Schizophrenia, Mood Disorders, Alzheimer's Disease, Mild Cognitive Impairment, Amyotrophic Lateral Sclerosis, Frontotemporal Dementia, Epilepsy and Parkinson's Disease). Written informed consent of the included healthy controls and/or patients was obtained by each of the participants at their local institute. 10 TB online storage space and 50,000+ CPU hours was reserved on the Cartesius supercomputer of the collaborative Information and Communication Technology (ICT) organization for Dutch education and research (SURF, <https://surfsara.nl/>) to analyze the data during the workshop. Workshop participants performed data quality checks on their data one week before the event after which they uploaded the MRI data (Diffusion Weighted Images (DWI) and pre-processed T1 data, see Materials and Methods) to their own user account on the supercomputer. During the workshop, participants were brought up to speed on DWI processing, connectome construction (see methods for details on the performed analysis), and running parallel jobs on a supercomputer. Together, a total of 15,947 MRI datasets were processed into anatomical connectome maps, with each output dataset including connectivity matrices with different types of connection weights and multiple

parcellation resolutions. Data processing was paralleled by interactive educational talks and workshops on connectome analysis.

Open data

In line with the collaborative nature of the event, the 10K group discussed making the connectome maps available to the scientific community for non-commercial use, free of restrictions. We include herein the resulting individual connectome maps of 8,000+ connectome datasets across an age range of 0 - 90 years, with five different edge weights (number of traced streamlines (NOS), streamline density (SD), fiber length, fractional anisotropy (FA) and mean diffusivity (MD)) at three parcellation resolutions (80+ cortical and subcortical regions, 100+ and 200+ cortical regions, see methods for details). Connectome maps are presented anonymously and blinded for participation site, together with basic demographics (age in bins of 5 years, gender, patient/control status, Fig. 1). Data is presented under the Non-Commercial Common Creative (CC BY-NC) license, free for all scientists to use in a non-commercial setting. A download request can be made at dutchconnectomelab.nl/10Kdata for a download link to the data (including 3 atlas resolutions, connectome matrices with multiple weights, information on the cortical and subcortical nodes, subject demographics (gender, age in 5 year bins, case/control)).

A first peek

We performed a few first analyses on the joint dataset, such as comparison to Human Connectome Project (HCP) data and examination of effects of age (see Materials and Methods for more detail). We observed a high consistency of the group averaged matrix with data derived from the high-quality HCP, with at least 69% of pathways identified in HCP also observed in the 10K set and with 98% of all non-existing connections in HCP verified in the 10K set. Furthermore, the distribution of weights across reconstructed connections is highly similar across the two datasets (FA weights, $r=.93$, $p<0.0001$, Fig. 1). Inverse MD showed rapid growth of microstructure in early years, with continuing development throughout adolescence, peaking around the beginning of the third decade, followed by a steady pattern of decline throughout aging (Fig. 1). Age analysis of the 10K set shows clear developmental patterns of cortical morphology (Fig. 1 showing cortical thickness) and white matter microstructure (Fig. 1 showing inverse MD) across age.

Future events

We acknowledge that there are many shortcomings to the presented MRI connectome dataset. Besides general, inherent limitations of diffusion MRI (Mori and Zhang, 2006), the presented dataset is a collation of data from a wide variety of groups, acquired with different scanners, different scanning protocols, varying data quality et cetera, as well as including data from a mixture of different patient and control populations. While these limitations place constraints on the type of investigations that one can perform with such collated multi-site datasets, we are

optimistic that the 10K dataset can be used as a large reference dataset for future studies, enabling many technical and neuroscientific research questions to be addressed (e.g. Fig. 1). As such, we hope that the presented data will be of use to the neuroscience community in the examination of the human connectome. Above all, we hope that our report will inspire others to organize exciting 10Kin1day-type of events in the near future, bringing together existing neuroimaging data and further catalyze open neuroimaging research of the healthy and diseased brain.

Author contributions

All authors contributed to the acquisition and/or analysis of the MRI data. Author Martijn van den Heuvel wrote the manuscript. All authors revised the manuscript and contributed intellectual content.

Acknowledgements

The 10Kin1day workshop was generously sponsored by the Neuroscience and Cognition program Utrecht (NCU) of the Utrecht University (<https://www.uu.nl/en/research/neuroscience-and-cognition-utrecht>), the ENIGMA consortium (<http://enigma.ini.usc.edu>), and personal grants: Martijn P. van den Heuvel: NWO-VIDI (452-16-015), MQ 2015 Fellowship; Simon Baron-Cohen: the Wellcome Trust; Medical Research Council UK; NIHR CLAHRC for Cambridgeshire and Peterborough Foundation National Health Services Trust; Autism Research Trust; Linda Booij: New Investigator Award, Canadian Institutes

of Health Research; Dara Cannon: Health Research Board (HRB), Ireland (grant code HRA-POR-2013-324); Sandra Chan: Research Grant Council (Hong Kong)-GRF 14101714; Eveline Crone: ERC-2010-StG-263234; Udo Dannlowski: DFG, grant FOR2107 DA1151/5-1, SFB-TRR58, Project C09, IZKF, grant Dan3/012/17; Stefan Du Plessis: MRC-RFA-UFSP-01-2013 ("Shared Roots" MRC Flagship grant); Thomas Frodl: Marie Curie Programme, International Training Programme, r'Birth; Dariusz Gąsecki: National Science Centre (UMO-2011/02/A/NZ5/00329); Beata Graff: National Science Centre (UMO-2011/02/A/NZ5/00329); Julie Hall: Western Sydney University Postgraduate Research Award; Laurena Holleran: Science Foundation Ireland, ERC; Helene Hopman: Research Grant Council (Hong Kong)-GRF 14101714; Lutz Jäncke: Velux Stiftung, grant 369; Andreas Jansen: DFG, grant FOR2107 JA 1890/7-1; Krzysztof Jodzio: National Science Centre (UMO-2013/09/N/HS6/02634); Vasiliy Kaleda: The Russian Foundation for Basic Research (grant code 15-06-05758 A); Tilo Kircher: DFG, grant FOR2107 KI 588/14-1, DFG, grant FOR2107 KI 588/15-1; Axel Krug: DFG, grant FOR2107 KO 4291/4-1, DFG, grant FOR2107 KO 4291/3-1; Irina Lebedeva: The Russian Foundation for Basic Research (grant code 15-06-05758 A); Edwin Lee: Health and Medical Research Fund - 11121271; Simon Lewis: NHMRC-ARC Dementia Fellowship 1110414, NHMRC Dementia Research Team Grant 1095127, NHMRC Project Grant 1062319; Carlos Lopez-Jaramillo: 537-2011, 2014-849; Andrew McIntosh: Wellcome Trust Strategic Award (104036/Z/14/Z); Christian Montag: Heisenberg-Grant, German Research Foundation, DFG MO 2363/3-1; Pedro Moreira: Foundation for

Science and Technology, Portugal - PDE/BDE/113601/2015; Krzysztof Narkiewicz: National Science Centre (UMO-2011/02/A/NZ5/00329); Patrycja Naumczyk: National Science Centre (UMO-2013/09/N/HS6/02634); Jiska Peper: NWO-Veni 451-10-007; Paul Rasser: PER and US would like to thank the Schizophrenia Research Institute and the Chief-Investigators of the Australian Schizophrenia Research Bank V. Carr, U. Schall, R. Scott, A. Jablensky, B. Mowry, P. Michie, S. Catts, F. Henskens, and C. Pantelis.; Agnieszka Sabisz: National Science Centre (UMO-2011/02/A/NZ5/00329); Stavros Skouras: European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 707730; Carles Soriano-Mas: Carlos III Health Institute (PI13/01958), Carlos III Health Institute (PI16/00889), Carlos III Health Institute (CPII16/00048); Edyta Szurowska: National Science Centre (UMO-2011/02/A/NZ5/00329); Alexander Tomyshov: The Russian Foundation for Basic Research (grant code 15-06-05758 A); Diana Tordesillas-Gutierrez: PI14/00918, PI14/00639; Leonardo Tozzi: Marie Curie Programme, International Training Programme, r'Birth; Sofie Valk: IMPRS Neurocom stipend; Theo van Erp: National Center for Research Resources at the National Institutes of Health (grant numbers: NIH 1 U24 RR021992 (Function Biomedical Informatics Research Network), NIH 1 U24 RR025736-01 (Biomedical Informatics Research Network Coordinating Center; <http://www.birncommunity.org>) and the NIH Big Data to Knowledge (BD2K) award (U54 EB020403 to Paul Thompson). Neeltje van Haren: NWO-VIDI (452-11-014); Judith van Leeuwen: NWO Veni (451-13-001); Marta Witkowska:

National Science Centre (UMO-2011/02/A/NZ5/00329); Veronica O'Keane: Meath Foundation. The funding sources had no role in the study design, data collection, analysis, and interpretation of the data.

HCP data was provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

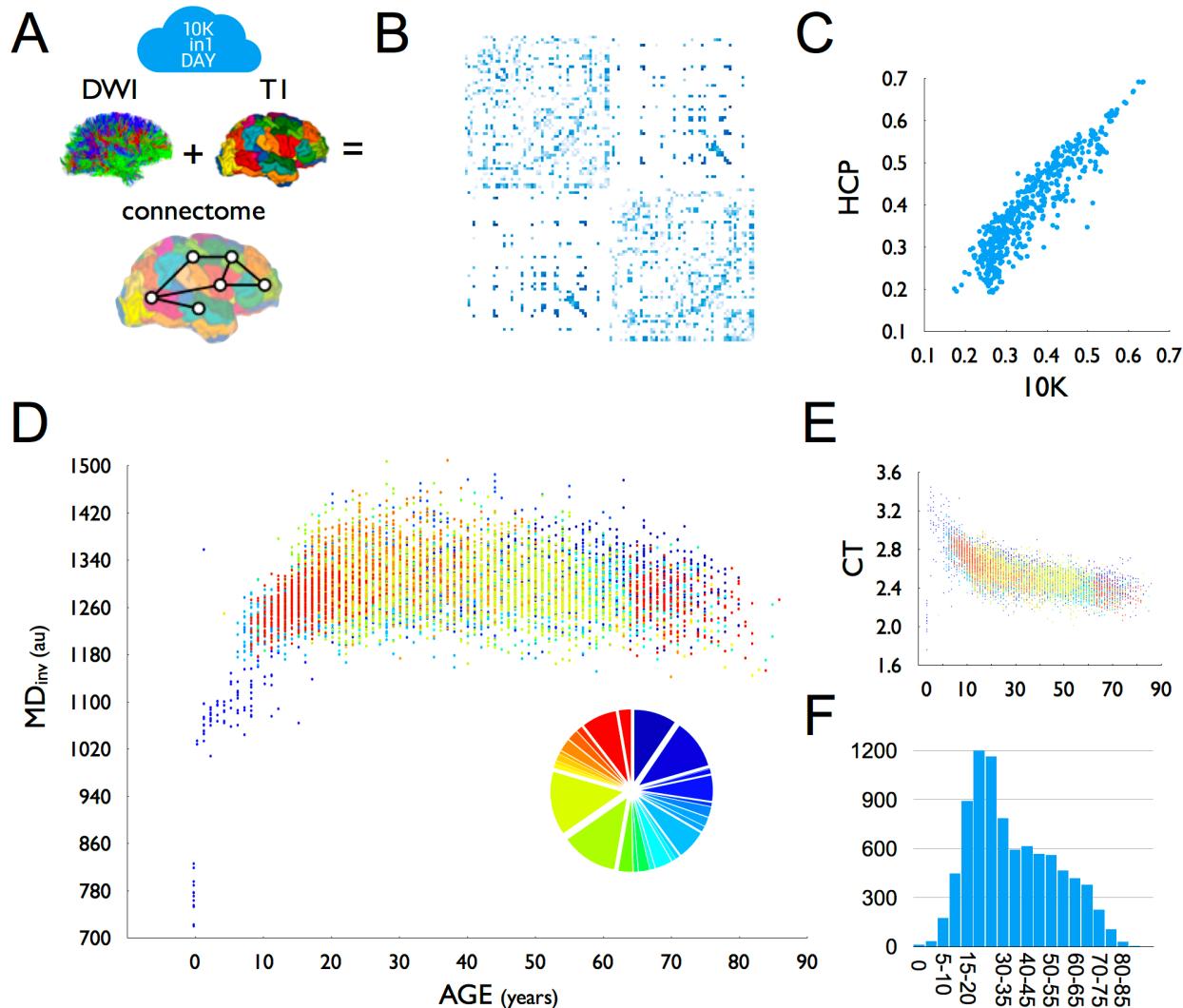


Figure 1. (A) For each dataset, DWI tractography was combined with T1-based parcellation of cerebral brain regions to reconstruct a brain network. (B) Group-averaged (group threshold 33%) FA matrix of the 10K dataset. (C) High overlap ($r=0.93$) between group-averaged FA values as derived from high-resolution HCP data and the 10K dataset. (D) Relationship between age and average inverse mean diffusivity (MD) across the 10K dataset. Colors indicate the different included datasets. Insert shows a pie diagram of the size of included datasets, color coded to set participation. One dataset (set_634413) was excluded from this plot, showing (across the age span)

deviating FA (lower) and MD (higher) values than the other datasets (see methods).

Due to the high total n, excluding this dataset did not change the relationship with age.

(E) Relationship between age and average cortical thickness (CT). (F) Age distribution of the presented data as in panel E and F. au=arbitrary units. DWI=diffusion weighted imaging. CT=cortical thickness.

Supplementary Files

- Supplementary Table 1 describing the dataset demographics
- Group files (3 atlas resolutions) containing 8,000+ connectome maps. A link to download the connectome matrices (2 Gb) can be obtained at <http://dutchconnectomelab.nl/10Kdata>.

Materials and Methods

A total of 42 groups (52 participants) participated in the workshop, some working on multiple datasets. Each dataset included a diffusion MRI scan and T1 MRI scan processed using FreeSurfer (Fischl and Dale, 2000). Datasets across groups included data from 1.5 and 3 Tesla MRI with varying scanner protocols and number of applied DWI gradients. Data included MRI data of healthy participants and patients with a neurological or psychiatric disorder. 24 groups were able to make their data available, making a total of 8,697 connectome maps publicly available through means of this report. Reconstructed connectome maps are presented anonymously, blinded for participation site and disease condition. Basic demographics of the datasets are included in the download set.

DWI Preprocessing. DWI datasets were corrected for susceptibility and eddy current distortions using the open tools from the FMRIB Software Library (FSL, <http://fsl.fmrib.ox.ac.uk>). Depending on the included DWI dataset, participants could choose to preprocess their data using the FSL *eddy_correct* or *eddy* tool (preprocessing scripts are included as supplementary information). For those DWI sets that included a subset of scans with an opposite k-space read out, an additional field distortion map could be formed and applied to the DWI images (Andersson et al., 2003).

Cortical parcellation. Before the event, the participants created FreeSurfer files based on their T1 images, with this output being subjected to varying degrees of quality control. The resulting parcellations of the cerebrum were used to select the regions of interest for the connectome reconstruction. The 68 cortical regions of FreeSurfer's standard Desikan-Killiany Atlas (Desikan et al., 2006, Fischl et al., 2004) as well as 14 subcortical regions were selected as network regions. Additionally, FreeSurfer files were used to further parcellate the cortex into 114 and 219 regions respectively using the Cammoun atlas (Cammoun et al., 2012).

Fiber reconstruction. After preprocessing of the DWI data, a diffusion tensor was fitted to the diffusion signal in each voxel of the white matter mask (selected based on the white matter segmentation map of the FreeSurfer files) using robust tensor fitting (Chang et al., 2005). Simple Diffusion Tensor Imaging (DTI) reconstruction was used due to its robustness and relatively low sensitivity to false positive reconstructions

compared to more advanced reconstruction methods (Klaus Maier-Hein, 2017), and thus potentially being the least distorting solution for connectome reconstruction and analysis based on MR imaging data (Zalesky et al., 2016). Decomposition of the tensor into eigenvectors and eigenvalues was used to select the main diffusion direction in each voxel, and to compute fractional anisotropy (FA) and mean diffusivity (MD) (Beaulieu and Allen, 1994). Deterministic fiber tractography was used to construct large-scale white matter pathways. Eight seeds (evenly distributed across the voxel) started in each white matter voxel, and fiber streamlines were formed by following the main diffusion direction from voxel to voxel using the *fiber assignment by continuous tracking* (FACT) algorithm (Mori and Barker, 1999), until one of the stopping criteria was met. A streamline was stopped when (1) it hit a voxel with an FA<0.1, (2) went out of the brain mask, or (3) made a turn >45 degrees.

Connectome reconstruction. A connectome map was made by combining the (sub)cortical parcellation map and the set of reconstructed fibers using commonly described procedures (Hagmann et al., 2008; van den Heuvel et al., 2012; van den Heuvel et al., 2010; van den Heuvel and Sporns, 2011). For each of the Cammoun Desikan-Killiany parcellation maps (i.e. 14+68, 14+114 and 14+219 regions respectively), the total collection of reconstructed fiber streamlines was used to assess the level of connectivity between each pair of (sub)cortical regions, represented as the *connectivity matrix CIJ*. (Sub)cortical regions were selected as the nodes of the reconstructed network, and for each combination of region *i* and region *j* where fiber streamlines touched both regions a connection

(i.e. network edge) was included in cell $C_{IJ}(i,j)$ in the connectivity matrix. Five different types of strength of a connection were computed and included as edge strength: (1) the number of reconstructed streamlines (NOS) between region i and j , (2) the average FA of the voxels traversed by the reconstructed streamlines, (3) the average MD of the reconstructed streamlines, (4) the average length of the reconstructed streamlines and (5) streamline density computed as the number of reconstructed streamlines corrected for the average volume of region i and region j (Hagmann et al., 2008).

Outliers. A total of 15,947 connectome maps were analyzed across the participating groups. Of the datasets that could be shared, 197 were detected as outliers (and were subsequently removed from the dataset). Outliers were detected automatically per group by testing for each connectome map their average connection strength and their distance to the group average prevalence map. The average connection strength of a connectome map was calculated for each of the five connection weights as the mean of the strengths over all existing (nonzero) connections. To measure the presence of odd connections or absence of common connections in a connectome map, we constructed a group prevalence matrix for each dataset, counting per node pair how many times an edge was observed across connectome maps in the group. For each connectome map the total prevalence of all observed connections and the total prevalence of all non observed connections was computed. Outliers were identified as connectome maps that displayed on any of the 7 measures (5

weight and 2 prevalence measures) a score below $Q1 - 2 \times IQR$ or above $Q3 + 2 \times IQR$, with $Q1$ and $Q3$ referring to the first and third quartile respectively and IQR the interquartile range $IQR = Q3 - Q1$. This resulted in the detection of 197 outliers in total, which were excluded from the dataset. One complete dataset (set_634413, $n=584$) showed across all included individual sets an average lower FA / higher MD as compared to the other datasets and this set was excluded from the age curves shown in Figure 1. Due to the high overall sample size, including or excluding this dataset did not change the shape of the final plot.

Comparison to HCP data

To test the validity of the 10K dataset, we compared the group average matrix of the 10K set to the group average matrix of data from the Human Connectome Project (HCP) (Van Essen et al., 2013). First, for the 10K dataset, a group average FA matrix was computed, by including those edges that were observed in at least 33% of the group (i.e. a group threshold of 33%, >2700 connectome maps showing a particular network edge). Average weight values of the included edges were taken as the non-zero mean of those edges across the connectome maps. Second, a similar group average FA matrix was derived from previously analyzed HCP data (van den Heuvel et al., 2016) ($n=487$ datasets). In brief, HCP analysis included the following steps (see (van den Heuvel et al., 2016) for more detailed information on the HCP data analysis). For each of the HCP DWI datasets a connectome was reconstructed based on the minimally pre-processed data of HCP. Given the high quality of the HCP data, analysis here included

reconstruction of multiple diffusion directions, allowing for the reconstruction of more complex fiber configurations (e.g. crossing fibers) (van den Heuvel et al., 2016). Similarly as for the 10K data, across the total set of 487 datasets, an average FA group matrix was computed, including those network edges that were observed in at least 33% of the total population (i.e. >160 datasets) and taking the non-zero mean of FA values across the group of subjects. Comparison between the 10K set and the HCP dataset was computed by means of (1) counting the number of existing connections and non-existing connections in the 10K dataset as observed in the HCP dataset and (2) by correlating the FA weights of the set of edges as observed in both datasets.

References

Amunts, K., Ebell, C., Muller, J., Telefont, M., Knoll, A., Lippert, T., 2016. The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain. *Neuron* 92, 574-581.

Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* 20, 870-888.

Beaulieu, C., Allen, P.S., 1994. Determinants of anisotropic water diffusion in nerves. *Magn Reson Med* 31, 394-400.

Cammoun, L., Gigandet, X., Meskaldji, D., Thiran, J.P., Sporns, O., Do, K.Q., Maeder, P., Meuli, R., Hagmann, P., 2012. Mapping the human connectome at multiple scales with diffusion spectrum MRI. *J Neurosci Methods* 203, 386-397.

Chang, L.C., Jones, D.K., Pierpaoli, C., 2005. RESTORE: robust estimation of tensors by outlier rejection. *Magn Reson Med* 53, 1088-1095.

Desikan et al., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage*, 31(3):968-80.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proc Natl Acad Sci U S A* 97, 11050-11055.

Fischl, B. et al., 2004. Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14, 11-22.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C.J., Wedeen, V.J., Sporns, O., 2008. Mapping the structural core of human cerebral cortex. *PLoS Biol* 6, e159.

Insel, T.R., Landis, S.C., Collins, F.S., 2013. Research priorities. The NIH BRAIN Initiative. *Science* 340, 687-688.

Jiang, T., 2013. Brainnetome: a new -ome to understand the brain and its disorders. *Neuroimage* 80, 263-272.

Klaus Maier-Hein, P.N., Jean-Christophe Houde, Marc-Alexandre Cote, Eleftherios Garyfallidis, Jidan Zhong, Maxime Chamberland, Fang-Cheng Yeh, Ying Chia Lin, Qing Ji, Wilburn E. Reddick, John O. Glass, David Qixiang Chen, Yuanjing Feng, Chengfeng Gao, Ye Wu, Jieyan Ma, He Renjie, Qiang Li, Carl-Fredrik Westin, Samuel Deslauriers-Gauthier, J. Omar Ocegueda Gonzalez, Michael Paquette, Samuel St-Jean, Gabriel Girard, Francois Rheault, Jasmeen Sidhu, Chantal M. W. Tax, Fenghua Guo, Hamed Y. Mesri, Szabolcs David, Martijn Froeling, Anneriet M. Heemskerk, Alexander Leemans, Arnaud Bore, Basile Pinsard, Christophe Bedetti, Matthieu Desrosiers, Simona Brambati, Julien Doyon, Alessia Sarica, Roberta Vasta, Antonio Cerasa, Aldo Quattrone, Jason Yeatman, Ali R. Khan, Wesley Hodges, Simon Alexander, Darcvi Romascano, Muhamed Barakovic, Anna Auria, Oscar Esteban, Alia Lemkadem, Jean-Philippe Thiran, H. Ertan Cetingul, Benjamin L. Odry, Boris Mailhe, Mariappan Nadar, Fabrizio Pizzagalli, Gautam Prasad, Julio Villalon-Reina, Justin Galvis, Paul Thompson, Francisco Requejo, Pedro Laguna, Luis Lacerda, Rachel Barrett, Flavio Dell'Acqua, Marco Catani, Laurent Petit, Emmanuel Caruyer, Alessandro Daducci, Tim Dyrby, Tim Holland-Letz, Claus Hilgetag, Bram Stieltjes, Maxime Descoteaux, 2017. Tractography-based connectomes are dominated by false-positive connections *bioRxiv*.

Mainen, Z.F., Hausser, M., Pouget, A., 2016. A better way to crack the brain. *Nature* 539, 159-161.

Mori, S., Barker, P.B., 1999. Diffusion magnetic resonance imaging: its principle and applications. *Anat Rec* 257, 102-109.

Mori, S., Zhang, J.Y., 2006. Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron* 51, 527-539.

Thompson, P.M., Andreassen, O.A., Arias-Vasquez, A., Bearden, C.E., Boedhoe, P.S., Brouwer, R.M., Buckner, R.L., Buitelaar, J.K., Bulayeva, K.B., Cannon, D.M., Cohen, R.A., Conrod, P.J., Dale, A.M., Deary, I.J., Dennis, E.L., de Reus, M.A., Desrivieres, S., Dima, D., Donohoe, G., Fisher, S.E., Fouche, J.P., Francks, C., Frangou, S., Franke, B., Ganjgahi, H., Garavan, H., Glahn, D.C., Grabe, H.J., Guadalupe, T., Gutman, B.A., Hashimoto, R., Hibar, D.P., Holland, D., Hoogman, M., Pol, H.E., Hosten, N., Jahanshad, N., Kelly, S., Kochunov, P., Kremen, W.S., Lee, P.H., Mackey, S., Martin, N.G., Mazoyer, B., McDonald, C., Medland, S.E., Morey, R.A., Nichols, T.E., Paus, T., Pausova, Z., Schmaal, L., Schumann, G., Shen, L., Sisodiya, S.M., Smit, D.J., Smoller, J.W., Stein, D.J., Stein, J.L., Toro, R., Turner, J.A., van den Heuvel, M.P., van den Heuvel, O.L., van Erp, T.G., van Rooij, D., Veltman, D.J., Walter, H., Wang, Y., Wardlaw, J.M., Whelan, C.D., Wright, M.J., Ye, J., Consortium, E., 2017. ENIGMA and the individual: Predicting factors that affect the brain in 35 countries worldwide. *Neuroimage* 145, 389-408.

van den Heuvel, M.P., Kahn, R.S., Goni, J., Sporns, O., 2012. High-cost, high-capacity backbone for global brain communication. *Proc Natl Acad Sci U S A* 109, 11372-11377.

van den Heuvel, M.P., Mandl, R.C., Stam, C.J., Kahn, R.S., Hulshoff Pol, H.E., 2010. Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis. *J Neurosci* 30, 15915-15926.

van den Heuvel, M.P., Scholtens, L.H., de Reus, M.A., Kahn, R.S., 2016. Associated Microscale Spine Density and Macroscale Connectivity Disruptions in Schizophrenia. *Biol Psychiatry* 80, 293-301.

van den Heuvel, M.P., Sporns, O., 2011. Rich-club organization of the human connectome. *J Neurosci* 31, 15775-15786.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., 2013. The WU-Minn Human Connectome Project: an overview. *Neuroimage* 80, 62-79.

Zalesky, A., Fornito, A., Cocchi, L., Gollo, L.L., van den Heuvel, M.P., Breakspear, M.,
2016. Connectome sensitivity or specificity: which is more important? *Neuroimage*.