

# brainlife.io: a decentralized and open-source cloud platform to support neuroscience research

Received: 10 March 2023

Accepted: 5 March 2024

Published online: 11 April 2024

 Check for updates

Soichi Hayashi<sup>1,30</sup>, Bradley A. Caron<sup>1,2,30</sup>, Anibal Sólón Heinsfeld<sup>2</sup>, Sophia Vinci-Booher<sup>1,3</sup>, Brent McPherson<sup>1,4</sup>, Daniel N. Bullock<sup>1</sup>, Giulia Bertò<sup>2</sup>, Guiomar Niso<sup>1,5</sup>, Sandra Hanekamp<sup>2</sup>, Daniel Levitas<sup>1,2</sup>, Kimberly Ray<sup>2</sup>, Anne MacKenzie<sup>2</sup>, Paolo Avesani<sup>6</sup>, Lindsey Kitchell<sup>1,7</sup>, Josiah K. Leong<sup>1,8</sup>, Filipi Nascimento-Silva<sup>1</sup>, Serge Koudoro<sup>1</sup>, Hanna Willis<sup>9</sup>, Jasleen K. Jolly<sup>10</sup>, Derek Pisner<sup>2</sup>, Taylor R. Zuidema<sup>1</sup>, Jan W. Kurzwski<sup>11</sup>, Kyriaki Mikellidou<sup>12,13</sup>, Aurore Bussalb<sup>14</sup>, Maximilien Chaumon<sup>14</sup>, Nathalie George<sup>14</sup>, Christopher Rorden<sup>15</sup>, Conner Victory<sup>16</sup>, Dheeraj Bhatia<sup>2</sup>, Dogu Baran Aydogan<sup>17,18</sup>, Fang-Cheng F. Yeh<sup>19</sup>, Franco Delogu<sup>16</sup>, Javier Guaje<sup>1</sup>, Jelle Veraart<sup>11</sup>, Jeremy Fischer<sup>1</sup>, Joshua Faskowitz<sup>1</sup>, Ricardo Fabrega<sup>1</sup>, David Hunt<sup>1</sup>, Shawn McKee<sup>20</sup>, Shawn T. Brown<sup>21</sup>, Stephanie Heyman<sup>22</sup>, Vittorio Iacovella<sup>23</sup>, Amanda F. Mejia<sup>1</sup>, Daniele Marinazzo<sup>12,24</sup>, R. Cameron Craddock<sup>2</sup>, Emanuele Olivetti<sup>23</sup>, Jamie L. Hanson<sup>19</sup>, Eleftherios Garyfallidis<sup>1</sup>, Dan Stanzione<sup>2</sup>, James Carson<sup>2</sup>, Robert Henschel<sup>1</sup>, David Y. Hancock<sup>1</sup>, Craig A. Stewart<sup>1</sup>, David Schnyer<sup>2</sup>, Damian O. Eke<sup>25</sup>, Russell A. Poldrack<sup>26</sup>, Steffen Bollmann<sup>27</sup>, Ashley Stewart<sup>27</sup>, Holly Bridge<sup>9</sup>, Ilaria Sani<sup>28,29</sup>, Winrich A. Freiwald<sup>28</sup>, Aina Puce<sup>1</sup>, Nicholas L. Port<sup>1</sup> & Franco Pestilli<sup>1,2</sup> ✉

Neuroscience is advancing standardization and tool development to support rigor and transparency. Consequently, data pipeline complexity has increased, hindering FAIR (findable, accessible, interoperable and reusable) access. brainlife.io was developed to democratize neuroimaging research. The platform provides data standardization, management, visualization and processing and automatically tracks the provenance history of thousands of data objects. Here, brainlife.io is described and evaluated for validity, reliability, reproducibility, replicability and scientific utility using four data modalities and 3,200 participants.

Over the past 30 years, neuroimaging has matured to adopt the FAIR (findable, accessible, interoperable and reusable) principles<sup>1,2</sup>, develop reporting best practices<sup>3</sup> and data standards<sup>4</sup>. While making research more rigorous and transparent, this maturation has inevitably increased compliance requirements. Indeed, just a few years ago it was possible to publish studies with a few hours of data collected and analyzed in a single laboratory. Today, studies require combining hundreds of

hours of measurement, across multiple participants, laboratories and data modalities (for example, magnetic resonance imaging (MRI), positron emission tomography, functional near-infrared spectroscopy, electro-encephalography (EEG) and magnetoencephalography (MEG)). To support the needs of a mature neuroimaging field, several data collection efforts have been started; relevant examples are the Human Connectome Project (HCP)<sup>5</sup>, Cambridge Centre for Ageing and

A full list of affiliations appears at the end of the paper. ✉ e-mail: [pestilli@utexas.edu](mailto:pestilli@utexas.edu)

Neuroscience study (Cam-CAN)<sup>6</sup>, Adolescent Brain Cognitive Development (ABCD) study<sup>7</sup>, the UK-Biobank<sup>8</sup>, Healthy Brain Network (HBN)<sup>9</sup>, Pediatric Imaging Neurocognition and Genetics (PING) study<sup>10</sup> and the Natural Scene Dataset<sup>11</sup>. At the same time, the complexity of the data pipeline has also increased with multiple, distinct, software libraries and analysis toolboxes developed<sup>12,13</sup>.

As compliance requirements grow, so do barriers to entry (Fig. 1a). The mature, neuroimaging field requires increased resources and technical training to piece together and track multiple processes such as data ingestion, standardization, storage, management, preprocessing and feature extraction (Fig. 1a). Currently, no single and low-barrier technology exists to integrate and manage the ever-changing software and data components of a full study. The growing compliance requirements affect the research community inequitably; smaller institutions and lower-income countries are more likely to lack resources and training. As such, this maturation process may risk favoring higher-resourced teams: an outcome that would not only decrease diversity and inclusion, but also slow-down scientific progress.

In support of simplicity, efficiency, transparency and equity in big data neuroscience research, our team has developed a community resource, brainlife.io (Fig. 1b). The brainlife.io platform stands on the pillars of open science (Fig. 1c), to provide free, secure and reproducible neuroscientific data analysis. Because of its web-based availability, brainlife.io should expand opportunities for researchers from nations and institutions with limited research budgets and resources. brainlife.io should then serve as an enabler for researchers and students from all sorts of institutions of higher education and all sorts of backgrounds to access cutting-edge neuroscience analytic tools.

brainlife.io is a ready-to-use and ready-to-expand platform. As a ready-to-use system, it allows researchers to upload and analyze data from MRI, MEG and EEG systems. Data are managed using a secure warehousing system with a proper access-control model. Data can be preprocessed and visualized using version-controlled applications (hereafter referred to as apps; <https://brainlife.io/apps>; Supplementary Fig. 1), compliant with major data standards (for example, the Brain Imaging Data Structure<sup>4</sup>). As a ready-to-expand system, software developers may submit apps guided by standardization and documentation (<https://github.com/brainlife/abcd-spec> and <https://brainlife.io/docs>). The platform uses opportunistic computing to serve commercial and academic clouds to researchers. Computing resources can be registered on brainlife.io for individual users and projects, or the larger community (Extended Data Fig. 1a,b). Supplementary Results 1 describe the technology.

The architecture of brainlife.io is based on a microservice approach for automated and decentralized data management and processing. Microservices are handled by the orchestration system Amaretti (Extended Data Fig. 1c,d and Extended Data Table 1) which deploys computational jobs on high-performance clusters and clouds (for example, Google Cloud, AWS or Microsoft Azure). Data management on brainlife.io is centered around projects, the 'one-stop-shop' for data management, processing, analysis and visualization (Supplementary Results 2 and Supplementary Fig. 2). Data archives can be docked by brainlife.io (Extended Data Fig. 1d) and data imported via the portal <https://brainlife.io/datasets> (Supplementary Table 1). Data from measurement instruments are imported using <https://brainlife.io/ezbid> (Extended Data Table 1)<sup>14</sup>. Data processing on brainlife.io uses an object-oriented and micro-workflow service model. Data objects are stored using predefined standardized formats, datatypes, that allow automated app pipelining (Extended Data Fig. 1e; <https://brainlife.io/datatypes>) and provenance tracking for millions of data objects. Data processing Apps are containerized, composable processing units, can be written in any language using containerization technology and are smart, meaning that they automatically identify, accept or reject data objects before processing (Supplementary Results 3 and

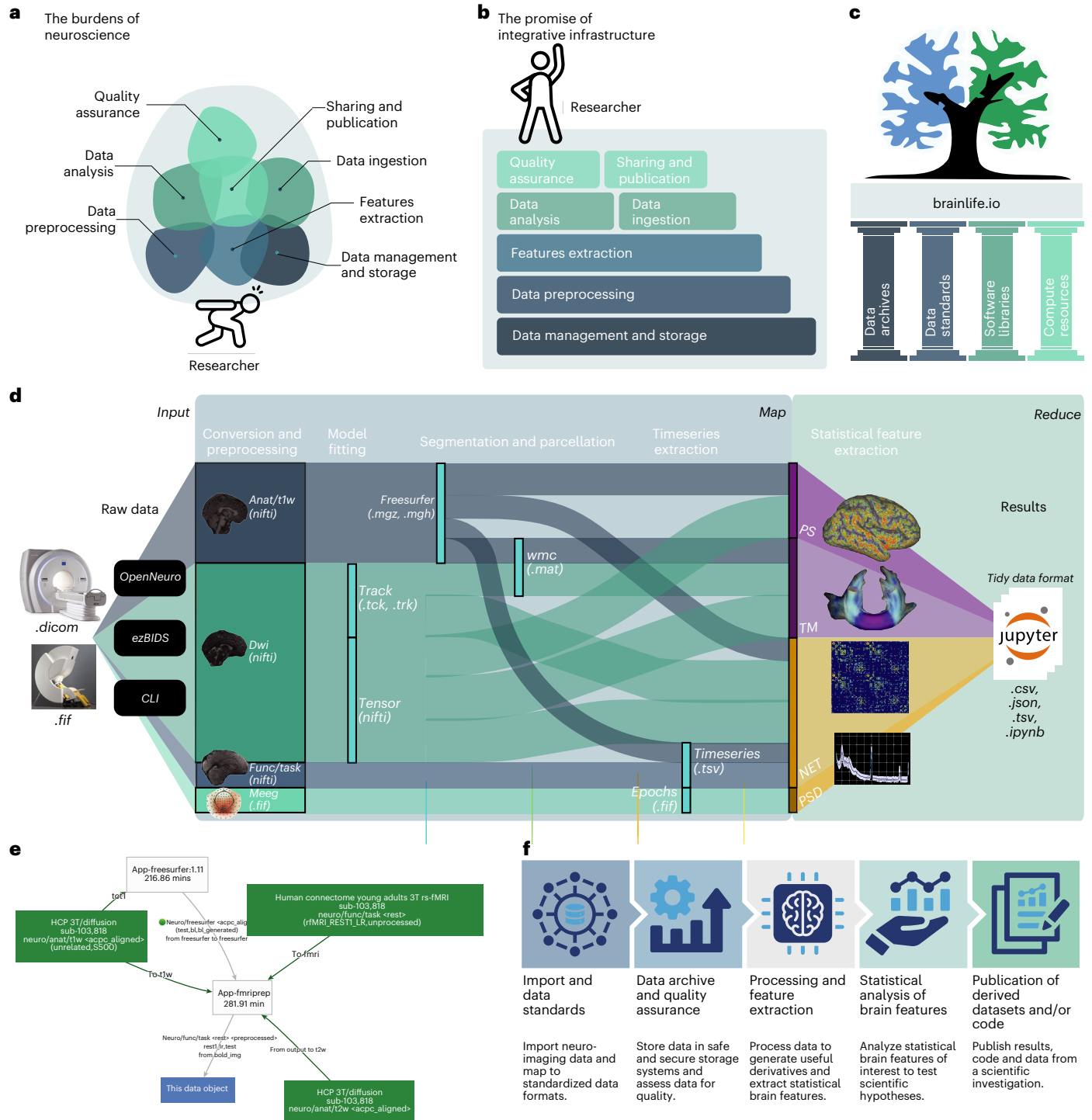
Supplementary Fig. 3). brainlife.io apps and datatypes are Brain Imaging Data Structure<sup>4</sup> compatible.

Complex neuroimaging processing pipelines are simplified into two main steps, akin to Google's MapReduce algorithm. An initial Map step preprocesses data objects asynchronously, in parallel, to extract features of interest (that is, functional activations, white matter maps, brain networks or time series data; Fig. 1d). A 'reduce' step follows where features extracted using apps are made available to preconfigured Jupyter Notebooks to perform analysis and generate figures. Indeed, all analyses and figures in this paper are available in brainlife.io notebooks (Supplementary Table 2). brainlife.io's data workflow makes it possible to integrate large volumes of data into small sets of features saved into 'tidy data' structures (Fig. 1d). For more documentation regarding usage of the platform, see Extended Data Fig. 2 and Supplementary Table 1. Datatypes inform apps allowing automated processing and provenance tracking for millions of data objects. brainlife.io tracks data object IDs, app versions and parameter sets across data processing steps. brainlife.io data provenance graphs visualize (Fig. 1e and Supplementary Table 1) and reproduce (Supplementary Table 1) the data generation steps. brainlife.io lowers the barriers of entry to FAIR neuroimaging by supporting an end-to-end data analysis workflow within a unified ecosystem (Fig. 1f).

We performed validation experiments to demonstrate cases where brainlife.io's technology produces results consistent with best practices in the field. We used over 1,800 participants from three datasets: PING, HCP<sub>s1200</sub> and Cam-CAN (Extended Data Figs. 3–8, Supplementary Results 4, Supplementary Fig. 4 and Supplementary Tables 3 and 4). Participants across all datasets spanned seven decades (that is, PING, 3–20 years; HCP<sub>s1200</sub>, 20–37 years and Cam-CAN, 18–88 years). Lifespan trajectories were plotted for multiple brain features (for example, brain region volume, white matter tract FAs, connectivity networks and MEG peak frequency; Fig. 2a and Extended Data Fig. 7) using brainlife.io's Jupyter Notebooks. Inverted U-shaped lifespan trajectories were estimated, consistent with previous studies<sup>15–17</sup> (Fig. 2a and Extended Data Fig. 7). The results generated using brainlife.io demonstrate that substantially different datasets can be collated to identify established brain's lifespan trajectories (Supplementary Results 4).

We further evaluated the ability to replicate results and generalize findings. Apps were created to estimate cortical thickness and tissue orientation dispersion, orientation dispersion index (ODI) and analyze the HCP<sub>s1200</sub> dataset. A negative relationship between cortical thickness and ODI was estimated (Fig. 2b and Extended Data Fig. 8;  $r_{\text{HCP-brainlife}} = -0.43$  versus  $r_{\text{original}}$ ), replicating the original study (ODI;  $r_{\text{original}} = -0.46$ )<sup>18</sup>. The result was also generalized to the Cam-CAN dataset (Fig. 2b and Extended Data Fig. 8;  $r_{\text{Cam-CAN-brainlife}} = -0.28$  versus  $r_{\text{original}}$ ). The association between life stressors and white matter organization of the uncinate fasciculus ( $r = -0.057$ ) was a replication of Hason et al.<sup>19</sup> using two independent datasets. The Negative Life Events Schedule (NLES) was correlated with quantitative anisotropy in the right- and left-hemisphere uncinate fasciculus (Fig. 2c and Extended Data Fig. 8;  $r_{\text{HBN\_LEFT}} = -0.35$ , two-tailed  $t$ -test,  $P = 0.018$ ;  $r_{\text{HBN\_RIGHT}} = -0.39$ , two-tailed  $t$ -test  $P < 0.0156$ ). Early Life Stress (a composite score of traumatic life events, environmental and neighborhood safety, and the family conflict subscale) was associated with the uncinate fasciculus FA (Fig. 2c and Extended Data Fig. 8;  $r_{\text{ABCD\_LEFT}} = -0.12$ ,  $P = 9.41 \times 10^{-5}$ ;  $r_{\text{ABCD\_RIGHT}} = -0.09$ ,  $P = 0.0035$ ). The results demonstrate the ability of brainlife.io services to detect meaningful associations in large, heterogeneous datasets (Supplementary Results 4).

Finally, we tested the ability of brainlife.io's services to detect optic radiation white matter changes as a result of eye disease<sup>20</sup>. Individuals with Stargardt's disease (deterioration initiated in the central retina) and choroideremia (deterioration initiated in peripheral retina) were compared to healthy controls. Stargardt's FA was reduced in optic radiation fibers projecting to central V1 (not peripheral; Fig. 2d).



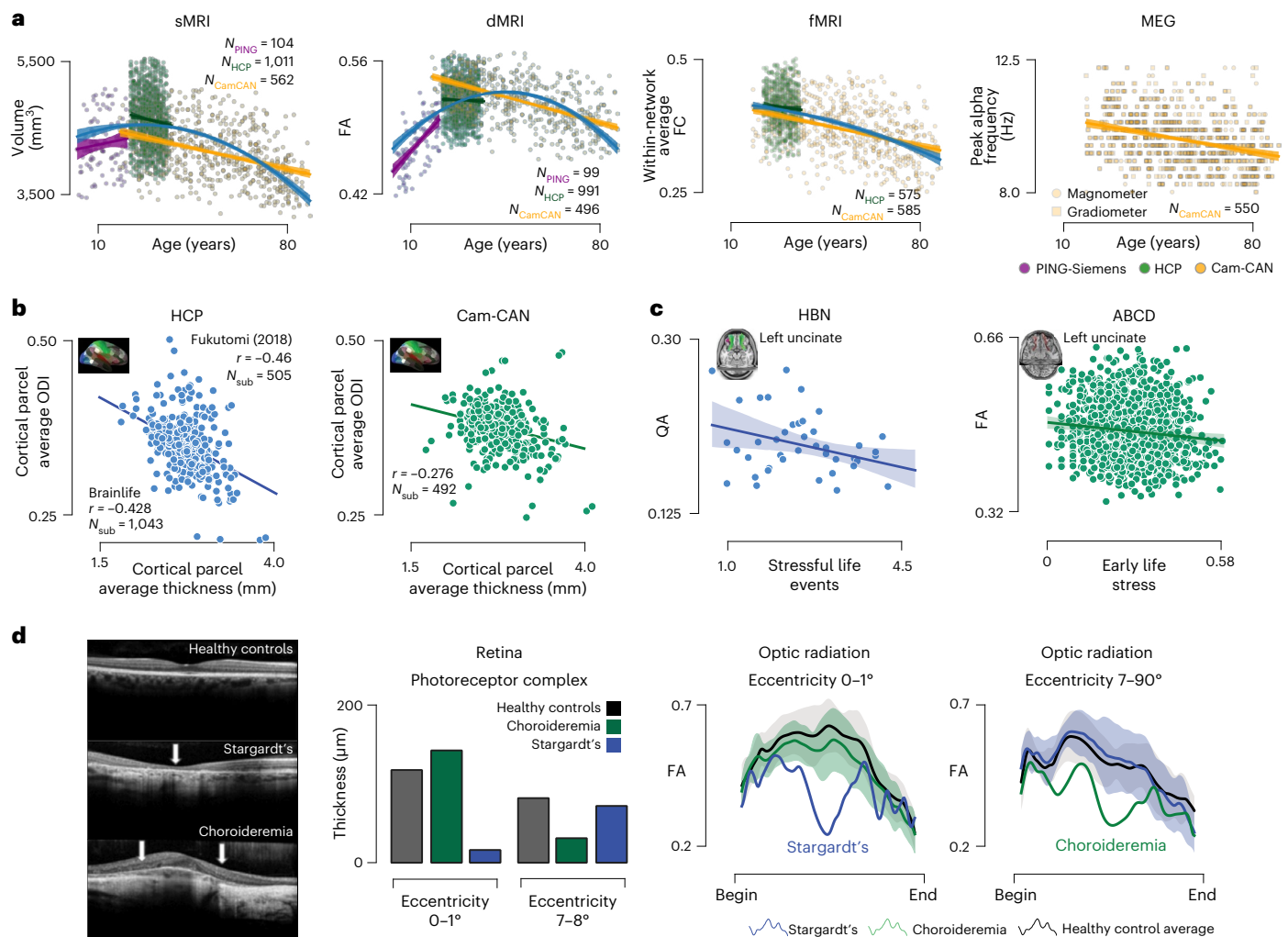
**Fig. 1 | The burdens of neuroscience and the promise of integrative infrastructure.** **a**, A figurative representation of the current major burdens of performing neuroimaging investigations. **b**, Our proposal for integrative infrastructure that coordinates services required to perform FAIR, reproducible, rigorous and transparent neuroimaging research thereby lifting the burden from the researcher. **c**, brainlife.io rests on the foundational pillars of the open science community such as data archives, standards, software libraries and compute resources. **d**, brainlife.io’s Map step takes MRI, MEG and EEG data and processes

them to extract statistical features of interest. brainlife.io’s reduce step takes the extracted features and serves them to Jupyter Notebooks for statistical analysis. PS, parc-stats datatype; TM, tractmeasures datatype; NET, network datatype and CLI, common line interface. **e**, The brainlife.io technology automates capture of data provenance. All data objects on brainlife.io are stored with a record of the apps, app versions and parameters used to process the data. **f**, The primary services are provided to the user by brainlife.io. Panels **a** and **b** adapted from ref. 22 under a Creative Commons license CC BY 4.0.

Choroideremia’s FA was reduced in optic radiation fibers projecting to peripheral V1 (not central; Fig. 2d and Supplementary Results 4).

Our vision for brainlife.io is that of a trusted, interoperable and integrative platform connecting data archives and global communities

of software developers, hardware providers and domain scientists (Supplementary Results 5 and Supplementary Table 5). The goal of brainlife.io is to facilitate research and education, accelerate brain understanding and lead to cures for brain diseases. To support this



**Fig. 2 | Brainlife.io supports scientific discovery and replication.**

**a–d**, Identifying unique relationships with brain features over the lifespan.

**a**, Relationship between participant age and right hippocampal volume, right inferior longitudinal fasciculus (FA, within-network average functional connectivity (FC) derived using the Yeo17 atlas and peak frequency in the alpha band derived from magnetometer (squares) and gradiometers (circles) from MEG data. These analyses include participants from the PING (purple), HCP<sub>1200</sub> (green) and Cam-CAN (yellow) datasets. Linear regressions were fitted to each dataset, and a quadratic regression was fitted to the entire dataset (blue).

**b,c**, Replication and generalization of previously reported scientific findings. **b**, Average cortical hcp-mmp parcel thickness ( $N_{\text{struc}} = 322$ ) compared to parcel the ODI from the NODDI model mapped to the cortical surface (inset) of the HCPS1200 dataset ( $N_{\text{sub}} = 1,043$ ) and Cam-CAN ( $N_{\text{sub}} = 492$ ) dataset compared to the parcel-average cortical thickness. **c**, Stressful life events were obtained from the NLES survey from HBN participants ( $N_{\text{sub}} = 42$ ) compared to uncinate-average

normalized quantitative anisotropy (QA). Mean linear regression (blue line)

fits and standard deviation (shaded blue). Early life stress was obtained from multiple surveys collected from ABCD participants ( $N_{\text{sub}} = 1,107$ ) compared to uncinate-average FA. Linear regression (green line) fits the data with standard deviation (shaded green). **d**, Identification of clinical biomarkers. **d**, Retinal optical coherence tomography images from healthy controls (top row), patients with Stargardt's disease (middle row) and patients with Choroideremia (bottom row). From these images, photoreceptor complex thickness was measured for each group (controls, gray; Choroideremia, green; Stargardt's, blue) in two distinct areas of the retina: the fovea (eccentricities 0–1°) and periphery (eccentricities 7–8°). In addition, optic radiations carrying information for each retinal area were segmented and FA profiles were mapped. Average profiles with standard error (shaded regions) were computed. One participant with Stargardt and one with Choroideremia were identified each having FA profiles that deviated from healthy controls.

vision, brainlife.io connects trainees, researchers, developers and computing resource managers in high-, medium- and low-income countries via technology. The platform is registered on fairsharing.org, datacite.org and nitric.org, it is recommended by the International Neuroinformatics Coordinating Facility (<https://incf.org/infrastructure/brainlife>) and it can serve the US National Institutes of Health in the United States data deposition and sharing mandate<sup>21,22</sup>. A comprehensive overview of the platform and tutorials can be found at <https://brainlife.io/docs>. Videos provide tutorials and demonstrations at [youtube.com/@brainlifeio](https://youtube.com/@brainlifeio). A slack channel supports communication and operations: <https://brainlife.slack.com>. Questions can be posted using the topic 'brainlife' on <https://neurostars.org> or GitHub issues

(<https://github.com/brainlife/brainlife/issues>) can be added directly to the code repositories. A quarterly outreach newsletter is sent out to all users, and an X account (@brainlifeio) informs the wider community about critical events. The platform has already collected a growing community (Supplementary Results 3).

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-024-02237-2>.



## References

1. Poldrack, R. A. et al. Scanning the horizon: towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* **18**, 115–126 (2017).
2. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
3. Nichols, T. E. et al. Best practices in data analysis and sharing in neuroimaging using MRI. *Nat. Neurosci.* **20**, 299–303 (2017).
4. Gorgolewski, K. J. et al. The Brain Imaging Data Structure, a format for organizing and describing outputs of neuroimaging experiments. *Sci. Data* **3**, 160044 (2016).
5. Van Essen, D. C. et al. The Human Connectome Project: a data acquisition perspective. *Neuroimage* **62**, 2222–2231 (2012).
6. Shafto, M. A. et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurol.* **14**, 204 (2014).
7. Casey, B. J. et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev. Cogn. Neurosci.* **32**, 43–54 (2018).
8. Sudlow, C. et al. UK BioBank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
9. Alexander, L. M. et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci. Data* **4**, 170181 (2017).
10. Jernigan, T. L. et al. The Pediatric Imaging, Neurocognition, and Genetics (PING) data repository. *Neuroimage* **124**, 1149–1154 (2016).
11. Allen, E. J. et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nat. Neurosci.* **25**, 116–126 (2022).
12. Markiewicz, C. J. et al. The OpenNeuro resource for sharing of neuroscience data. *eLife* **10**, e71774 (2021).
13. Poldrack, R. A., Gorgolewski, K. J. & Varoquaux, G. Computational and informatic advances for reproducible data analysis in neuroimaging. *Annu. Rev. Biomed. Data Sci.* <https://doi.org/10.1146/annurev-biodatasci-072018-021237> (2019).
14. Levitas, D. et al. ezBIDS: guided standardization of neuroimaging data interoperable with major data archives and platforms. *Sci. Data* **11**, 179 (2024).
15. Betzel, R. F. et al. Changes in structural and functional connectivity among resting-state networks across the human lifespan. *Neuroimage* **102**, 345–357 (2014).
16. Bethlehem, R. A. I. et al. Brain charts for the human lifespan. *Nature* **604**, 525–533 (2022).
17. Yeatman, J. D., Wandell, B. A. & Mezer, A. A. Lifespan maturation and degeneration of human brain white matter. *Nat. Commun.* **5**, 4932 (2014).
18. Fukutomi, H. et al. Neurite imaging reveals microstructural variations in human cerebral cortical gray matter. *Neuroimage* **182**, 488–499 (2018).
19. Hanson, J. L., Knodt, A. R., Brigidi, B. D. & Hariri, A. R. Lower structural integrity of the uncinate fasciculus is associated with a history of child maltreatment and future psychological vulnerability to stress. *Dev. Psychopathol.* **27**, 1611–1619 (2015).
20. Ogawa, S. et al. White matter consequences of retinal receptor and ganglion cell damage. *Invest. Ophthalmol. Vis. Sci.* **55**, 6976–6986 (2014).
21. Kozlov, M. NIH issues a seismic mandate: share data publicly. *Nature* <https://doi.org/10.1038/d41586-022-00402-1> (2022).
22. Eke, D. O. et al. International data governance for neuroscience. *Neuron* <https://doi.org/10.1016/j.neuron.2021.11.017> (2021).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024, corrected publication 2024

<sup>1</sup>Indiana University, Bloomington, IN, USA. <sup>2</sup>The University of Texas, Austin, TX, USA. <sup>3</sup>Vanderbilt University, Nashville, TN, USA. <sup>4</sup>McGill University, Montréal, Quebec, Canada. <sup>5</sup>Cajal Institute, CSIC, Madrid, Spain. <sup>6</sup>Fondazione Bruno Kessler, Trento, Italy. <sup>7</sup>Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, USA. <sup>8</sup>University of Arkansas, Fayetteville, AR, USA. <sup>9</sup>University of Oxford, Headington, Oxford, UK. <sup>10</sup>Anglia Ruskin University, Cambridge, UK. <sup>11</sup>New York University, New York, NY, USA. <sup>12</sup>University of Limassol, Nicosia, Cyprus. <sup>13</sup>University of Cyprus, Nicosia, Cyprus. <sup>14</sup>Institut du Cerveau, CNRS, Sorbonne Université, Paris, France. <sup>15</sup>University of South Carolina, Columbia, SC, USA. <sup>16</sup>Lawrence Technological University, Southfield, MI, USA. <sup>17</sup>University of Eastern Finland, Kuopio, Finland. <sup>18</sup>Aalto University School of Science, Espoo, Finland. <sup>19</sup>University of Pittsburgh, Pittsburgh, PA, USA. <sup>20</sup>University of Michigan, Ann Arbor, MI, USA. <sup>21</sup>Hewlett-Packard Enterprise, Pittsburgh, PA, USA. <sup>22</sup>SHEGEL, Massul, Luxembourg. <sup>23</sup>University of Trento, Rovereto, Italy. <sup>24</sup>University of Ghent, Ghent, Belgium. <sup>25</sup>University of Nottingham, Nottingham, UK. <sup>26</sup>Stanford University, Stanford, CA, USA. <sup>27</sup>University of Queensland, St Lucia, Queensland, Australia. <sup>28</sup>The Rockefeller University, New York, NY, USA. <sup>29</sup>University of Geneva, Geneva, Switzerland. <sup>30</sup>These authors contributed equally: Soichi Hayashi, Bradley A. Caron. ✉e-mail: [pestilli@utexas.edu](mailto:pestilli@utexas.edu)

## Methods

### Data sources

Multiple openly available data sources were used for examining the validity, reliability and reproducibility of brainlife.io apps and for examining population distributions. All information regarding the specific image acquisitions, participant demographics and study-wide preprocessing can be found in the publications in refs. 5–7,9,10,23–25. Some data sources are currently unpublished. For these, the appropriate information is provided. Experiments were approved by the local institutional review boards (IRB) and only the personnel approved for a specific study accessed the data in private projects on brainlife.io.

### Validity, reliability, reproducibility, replicability, developmental trends and reference datasets

**HCP (test–retest, s1200-release).** Data from these projects were used to assess the validity, reliability and reproducibility of the platform. They were used to assess the abilities of the platform to identify developmental trends in structural and functional measures, and they were used to generate reference datasets. For structural MRI (sMRI) data, the minimally preprocessed structural T1w and T2w images from the HCP from 1,066 participants from the s1200 and 44 participants from the test–retest releases were used<sup>5</sup>. Specifically, the 1.25 mm ‘acpc\_dc\_restored’ images generated from the Siemens 3 T MRI scanner were used for all analyses involving the HCP. For most examinations, the already-processed Freesurfer output from HCP was used. For diffusion MRI (dMRI) data, to assess the validity of preprocessing on brainlife.io, the unprocessed dMRI data from 44 participants from the HCP test dataset was used. For reliability and all remaining analyses, the minimally preprocessed dMRI images from 1,066 participants from the s1200 and 44 participants from the test–retest releases from the 3 T Siemens scanner were used. All processes incorporated the multi-shell acquisition data. For functional data (functional MRI (fMRI)), regarding validation, the unprocessed resting-state fMRI data from 44 participants from the HCP test dataset were compared to the minimally preprocessed blood oxygenation level dependent data provided by HCP. For reliability and all other analyses, the minimally preprocessed blood oxygenation level dependent data from 1,066 participants from the s1200 and 44 participants from the test–retest releases from the 3 T Siemens scanner were used.

**The Cam-CAN.** The data from this project were used to assess the validity, reliability and reproducibility of the platform, and to assess the abilities of the platform to identify developmental trends of structural and functional measures, and to generate reference datasets. For sMRI data, the unprocessed 1 mm isotropic structural T1w and T2w images from 652 participants from the Cam-CAN<sup>6</sup> study were used. For dMRI data, the unprocessed 2 mm isotropic diffusion (dMRI) images from 652 participants from the Cam-CAN study were used. For fMRI data, the  $3 \times 3 \times 4$  mm<sup>3</sup> unprocessed resting-state fMRI images from 652 participants from the Cam-CAN study were used. For electromagnetic data (MEG), the 1,000 Hz resting-state filtered and unfiltered datasets from 652 participants from the Cam-CAN study were used.

### Developmental trends and reference datasets

**PING.** The data from this project were used to assess the abilities of the platform to identify developmental trends of structural measures and to generate reference datasets. For sMRI data, the unprocessed  $1.2 \times 1.0 \times 1.0$  mm<sup>3</sup> structural T1w and the 1.0 mm isotropic T2w images from 110 participants from the PING<sup>10</sup> study were used. For dMRI data, the unprocessed 2 mm isotropic diffusion (dMRI) images from 110 participants from the PING study were used.

### Replicability datasets

**ABCD.** For sMRI data, the unprocessed 1 mm isotropic structural T1w and T2w images from a subset of 1,877 participants from the

ABCD (release-2.0.0) study were used. For dMRI data, the unprocessed 1.77 mm isotropic diffusion (dMRI) images from a subset of 1,877 participants from the ABCD (release-2.0.0) study were used<sup>7,26</sup>. A single diffusion gradient shell was used for these experiments ( $b = 3,000$  s mm<sup>-2</sup>). Research was approved by the University of Arkansas IRB (no. 2209425822).

**HBN.** The data from this project were used to assess the abilities of the platform to replicate previously published findings via the assessment of the relationship between microstructural measures mapped to segmented uncinate fasciculi and self-reported early life stressors. Research was approved by the University of Pittsburgh IRB (no. PRO17060350). For sMRI data, the 0.8 mm isotropic structural T1w images from 42 participants from the HBN study<sup>9</sup> were used. For dMRI data, the unprocessed 1.8 mm isotropic diffusion (dMRI) images from 42 participants from the CitiGroup Cornell Brain Imaging Center site of the HBN study were used. Research was approved by the University of Pittsburgh IRB (no. PRO17060350).

**UPENN-PMC.** The University of Pennsylvania, Penn Memory Center (UPENN-PMC) data from this project were used to assess the abilities of the platform to replicate previously published findings via the assessment of the performance of an automated hippocampal segmentation algorithm. Secondary data analyses were conducted under IRB exemption at Indiana University. For sMRI data, the T1w and T2w data were provided within the Automated Segmentation of Hippocampal Subfields Automated Segmentation of Hippocampal Subfields atlas<sup>27</sup>.

### Clinical-identification datasets

**Indiana University Acute Concussion dataset.** The data from this project were used to assess the abilities of the platform to identify clinical populations via the mapping of microstructural measures to the cortical surface. Neuroimaging was performed at the Indiana University Imaging Research Facility, housed within the Department of Psychological and Brain Sciences with a 3 T Siemens Prisma whole-body MRI using a 64-channel head coil. Within this study, nine concussed athletes and 20 healthy athletes were included. Research approved by Indiana University (IRB 906000405). For sMRI data, high-resolution T1-weighted structural volumes were acquired using an MPRAGE sequence: TI = 900 ms, TE = 2.7 ms, TR = 1,800 ms, flip angle 9°, with 192 sagittal slices of 1.0 mm thickness, a field of view of  $256 \times 256$  mm<sup>2</sup> and an isometric voxel size of 1.0 mm<sup>3</sup> (where TI, TE and TR refer to inversion time, echo time and repetition time, respectively). The total acquisition time was 4 min and 34 s. High-resolution T2-weighted structural volumes were also acquired: TE = 564 ms, TR = 3,200 ms, flip angle 120°, with 192 sagittal slices, a field of view of  $240 \times 256$  mm<sup>2</sup> and an isometric voxel size of 1.0 mm<sup>3</sup>. Total acquisition time was 4 min and 30 s. Diffusion data (dMRI) were collected using single-shot spin-echo simultaneous multi-slice (SMS) echo-planar imaging (transverse orientation, TE = 92.00 ms, TR = 3,820 ms, flip angle 78°, isotropic 1.5 mm<sup>3</sup> resolution; FOV = LR 228 × 228 × 144 mm<sup>3</sup>; acquisition matrix MxP 138 × 138. SMS acceleration factor 4). This sequence was collected twice, one in the anterior-posterior fold-over direction and the other in the posterior-anterior (PA) fold-over direction, with the same diffusion gradient strengths and the number of diffusion directions: 30 diffusion directions at  $b = 1,000$  s mm<sup>-2</sup>, 60 diffusion directions at  $b = 1,750$  s mm<sup>-2</sup>, 90 diffusion directions at  $b = 2,500$  s mm<sup>-2</sup> and 19  $b = 0$  s mm<sup>-2</sup> volumes. The total acquisition time for both sets of dMRI sequences was 25 min and 58 s.

### Oxford University Choroideremia & Stargardt’s Disease Dataset.

The data from this project were used to assess the abilities of the platform to identify clinical populations via mapping retinal-layer thickness via optical coherence tomography and mapping of microstructural measures along optic radiation bundles segmented using visual field

information (eccentricity). Neuroimaging was performed at the Wellcome Centre for Integrative Neuroimaging, Oxford with the Siemens 3 T scanner. Research was approved by the UK Health Regulatory Authority reference 17/LO/1540. For sMRI data, high-resolution T1-weighted anatomical volumes were acquired using an MPRAGE sequence: TI = 904 ms, TE = 3.97 ms, TR = 1,900 ms, flip angle 8°, with 192 sagittal slices of 1.0 mm thickness, a field of view of  $174 \times 192 \times 192 \text{ mm}^3$  and an isometric voxel size of  $1.0 \text{ mm}^3$ . The total acquisition time was 5 min and 31 s. Diffusion data (dMRI) were collected using echo-planar imaging (transverse orientation, TE = 92.00 ms, TR = 3,600 ms, flip angle 78°,  $2.019 \times 2.019 \times 2.0 \text{ mm}^3$  resolution; FOV =  $210 \times 220 \times 158 \text{ mm}^3$ ; acquisition matrix  $M \times P = 210 \times 210$ , SMS acceleration factor 3). This sequence was collected twice, one in the anterior-posterior fold-over direction and the other in the PA fold-over direction. The PA fold-over scan contained six diffusion directions, three at  $b = 0 \text{ s mm}^{-2}$  and three at  $b = 2,000 \text{ s mm}^{-2}$ , and was used primarily for susceptibility-weighted corrections. The anterior-posterior fold-over scan contained 105 diffusion directions, five at  $b = 0 \text{ mm s}^{-2}$ , 51 at  $b = 1,000 \text{ mm s}^{-2}$  and 49 at  $b = 2,000 \text{ mm s}^{-2}$ . The total acquisition time for both sets of dMRI sequences was 7 min and 8 s.

### General processing pipelines

**Structural processing.** For the ABCD, Cam-CAN, Oxford University Choroideremia & Stargardt's Disease Dataset, and the Indiana University Acute Concussion datasets, the structural T1w and T2w (sMRI) images (if available) were preprocessed, including bias correction and alignment to the anterior commissure-posterior commissure plane, using the brainlife.io apps A273 (<https://doi.org/10.25663/brainlife.app.273>) and A350 (<https://doi.org/10.25663/brainlife.app.350>), respectively. For PING data, no bias correction was performed but alignment to the anterior commissure-posterior commissure plane was performed using A99 (<https://doi.org/10.25663/brainlife.app.99>) and A116 (<https://doi.org/10.25663/brainlife.app.116>) for T1w and T2w data, respectively. For HCP data, this data was already provided. The structural T<sub>1</sub>-weighted images for each participant and dataset were then segmented into different tissue types using functionality provided by MRTrix3 (ref. 28) implemented as A239 (<https://doi.org/10.25663/brainlife.app.239>). For a subset of datasets, this was performed within the diffusion tractography generation step using A319 (<https://doi.org/10.25663/brainlife.app.319>). The gray- and white-matter interface mask was subsequently used as a seed mask for white matter tractography. The processed structural T1w and T2w images were then used for segmentation and surface generation using the recon-all function from Freesurfer<sup>29</sup> (AO; <https://doi.org/10.25663/brainlife.app.0>). Following Freesurfer, representations of the cortical 'midthickness' surface were computed by spatially averaging the coordinates of the pial and white matter surfaces generated by Freesurfer using the wb\_command-surface-cortex-layer function provided by Workbench command for the HCP<sub>TR</sub>, HCP<sub>s1200</sub>, ABCD, Cam-CAN, PING and Indiana University Acute Concussion datasets. These surfaces were used for cortical tissue mapping analyses. Following Freesurfer and midthickness-surface generation, the 180 multimodal cortical nodes (hcp-mmp) atlas and the Yeo 17 (yeo17) atlas were mapped to the Freesurfer segmentation of each participant implemented as brainlife.io app A23 (<https://doi.org/10.25663/brainlife.app.23>). These parcellations were used for subsequent cortical, subcortical and network analyses. In addition, measures for cortical thickness, surface area, volume and summaries of diffusion models of microstructure were estimated using A383 (<https://doi.org/10.25663/brainlife.app.383>) and A389 (<https://doi.org/10.25663/brainlife.app.389>). To estimate population receptive fields and visual field eccentricity properties in the cortical surface in the Oxford University Choroideremia & Stargardt's Disease Dataset, the automated mapping algorithm developed by refs. 30,31 was implemented using A187 (<https://doi.org/10.25663/brainlife.app.187>). To segment thalamic nuclei for optic radiation tracking, the automated

thalamic nuclei segmentation algorithm provided by Freesurfer<sup>28</sup> was implemented as A222 (<https://doi.org/10.25663/brainlife.app.222>). Finally, visual regions of interest (ROI) binned by eccentricity were then generated using AFNI software<sup>32</sup> functions implemented in A414 (<https://doi.org/10.25663/brainlife.app.414>). To assess the replicability capabilities of the platform, an automated hippocampal nuclei segmentation app (A262; <https://doi.org/10.25663/brainlife.app.262>) was used to segment hippocampal subfields from participants within the UPENN-PMC dataset provided within the Automated Segmentation of Hippocampal Subfields atlas.

### dMRI processing

**Preprocessing and model fitting.** For most of the analyses involving the HCP dataset, the minimally preprocessed dMRI images were used and thus no further preprocessing was performed. However, to assess the validity of the preprocessing pipeline, the unprocessed dMRI data from the HCP test dataset and dMRI images were preprocessed following the protocol outlined in ref. 33 using A68 (<https://doi.org/10.25663/brainlife.app.68>). The same app was also used for preprocessing the dMRI images for the ABCD, Cam-CAN, PING, Oxford University Choroideremia & Stargardt's Disease Dataset, the Indiana University Acute Concussion and HBN datasets. Specifically, dMRI images were denoised and cleaned from Gibbs ringing using functionality provided by MRTrix3 before being corrected for susceptibility, motion and eddy distortions and artifacts using FSL's topup and eddy functions<sup>34,35</sup>. Eddy-current and motion correction was applied via the eddy\_cuda8.0 with the replacement of outlier slices (that is, repol) command provided by FSL<sup>36-39</sup>. Following these corrections, MRTrix3's dwigradcheck functionality was used to check and correct for potential misaligned gradient vectors following topup and eddy<sup>40</sup>. Next, dMRI images were debiased using ANT's n4 functionality<sup>41</sup> and the background noise was cleaned using MRTrix3.0's dwidenoise functionality<sup>42</sup>. Finally, the preprocessed dMRI images were registered to the structural (T1w) image using FSL's epi\_reg functionality<sup>43-45</sup>. Following preprocessing, brain masks for dMRI data using bet from FSL were implemented as A163 (<https://doi.org/10.25663/brainlife.app.163>).

**DTI, NODDI and  $q$ -sampling model fitting.** Following preprocessing, the diffusion tensor imaging (DTI) model<sup>46</sup> and the neurite orientation dispersion and density imaging (NODDI)<sup>47,48</sup> models were subsequently fit to the preprocessed dMRI images for each participant using either A319 (<https://doi.org/10.25663/brainlife.app.319>) or A292 (<https://doi.org/10.25663/brainlife.app.292>) for DTI model fitting and A365 (<https://doi.org/10.25663/brainlife.app.365>) for NODDI fitting. Note, the NODDI model was only fit on the HCP, Cam-CAN, Oxford University Choroideremia & Stargardt's Disease Dataset and the Indiana University Acute Concussion datasets. For those datasets, the NODDI model was fit using an intrinsic free diffusivity parameter ( $d_{\parallel}$ ) of  $1.7 \times 10^{-3} \text{ mm}^2 \text{ s}^{-1}$  for white matter tract and network analyses, and a  $d_{\parallel}$  of  $1.1 \times 10^{-3} \text{ mm}^2 \text{ s}^{-1}$  for cortical tissue mapping analyses, using AMICO's implementation<sup>48</sup> as A365 (<https://doi.org/10.25663/brainlife.app.365>). The constrained spherical deconvolution<sup>49</sup> model was then fit to the preprocessed dMRI data for each run across four spherical harmonic orders (that is,  $L_{\max}$ ) parameters (2, 4, 6, 8) using functionality provided by MRTrix3 implemented as brainlife.io app A238 (<https://doi.org/10.25663/brainlife.app.238>). For the PING datasets, the constrained spherical deconvolution model was fit using the same code found in A238 (<https://doi.org/10.25663/brainlife.app.238>), but performed using the tractography app A319 (<https://doi.org/10.25663/brainlife.app.319>). For the HBN dataset, the isotropic spin distribution function was obtained by reconstructing the diffusion MRI data with the generalized  $q$ -sampling imaging method<sup>50</sup> using functionality provided by DSI-Studio<sup>51</sup> (A423; <https://doi.org/10.25663/brainlife.app.423>). Quantitative anisotropy was then estimated from the isotropic spin distribution function.



**Tractography.** Following model fitting, the fiber orientation distribution functions for  $L_{\max} = 6$  and  $L_{\max} = 8$  were subsequently used to guide anatomically constrained probabilistic tractography<sup>52</sup> using functions provided by MRTrix3 implemented as brainlife.io app A297 (<https://doi.org/10.25663/brainlife.app.297>) or A319 (<https://doi.org/10.25663/brainlife.app.319>). For the HCP<sub>TR</sub>, HCP<sub>s1200</sub> and Oxford University Choroideremia & Stargardt's Disease datasets,  $L_{\max} = 8$  was used. For the ABCD and Cam-CAN datasets,  $L_{\max} = 6$  was used. For the HCP, ABCD and Cam-CAN, datasets, a total of 3 million streamlines were generated. For all datasets, a step size of 0.2 mm was implemented. For the HCP<sub>TR</sub>, HCP<sub>s1200</sub>, ABCD and Cam-CAN datasets, minimum and maximum lengths of streamlines were set at 25 and 250 mm, respectively, and a maximum angle of curvature of 35° was used. For the PING dataset, minimum and maximum lengths of streamlines were set at 20 and 220 mm, respectively, and a maximum angle of curvature of 35° was used.

**Whiter matter segmentation and cleaning.** Following tractography, 61 major white matter tracts were segmented for each run using a customized version of the white matter query language<sup>53</sup> implemented as brainlife.io app A188 (<https://doi.org/10.25663/brainlife.app.188>). Outlier streamlines were subsequently removed using functionality provided by Vistasoft and implemented as brainlife.io app A195 (<https://doi.org/10.25663/brainlife.app.195>). Following cleaning, tract profiles with 200 nodes were generated for all DTI and NODDI measures across the 61 tracts for each participant and test–retest condition using functionality provided by Vistasoft and implemented as A361 (<https://doi.org/10.25663/brainlife.app.361>). Macrostructural statistics, including average tract length, tract volume and streamline count were computed using functionality provided by Vistasoft implemented as A189 (<https://doi.org/10.25663/brainlife.app.189>). Microstructural and macrostructural statistics were then compiled into a single data frame using A397 (<https://doi.org/10.25663/brainlife.app.397>).

**Segmentation of the optic radiation.** To generate optic radiations segmented by estimates of visual field eccentricity in the Oxford University Choroideremia & Stargardt's Disease Dataset, ConTrack<sup>54</sup> tracking was implemented as A252 (<https://doi.org/10.25663/brainlife.app.252>). Then, 500,000 sample streamlines were generated using a step size of 1 mm. Samples were then pruned using inclusion and exclusion waypoint ROI following methodologies outlined in refs. 19,55.

**Segmentation of uncinate fasciculus.** To assess the relationship between uncinate tract-average quantitative anisotropy, fractional anisotropy (FA) and early life stressors within two independent datasets (HBN, ABCD), the tract-average quantitative anisotropy for the left and right uncinate were computed from 42 participants from the HBN and the tract-average FA were computed from 1,107 participants from the ABCD dataset. For the HBN dataset, a full tractography segmentation pipeline was used to preprocess the dMRI data and segment the uncinate fasciculus using A423 (<https://doi.org/10.25663/brainlife.app.423>). Automatic fiber tracking was then performed to segment the uncinate fasciculus using default parameters and templates from a population tractography atlas from the HCP<sup>56</sup>. A threshold of 16 mm as the maximum allowed threshold for the shortest streamline distance was then applied to remove spurious streamlines. The whole tract-average quantitative anisotropy was then estimated. To probe stress exposure within the HBN dataset, we used the NLES, a 22-item questionnaire in which participants were asked about the occurrence of different stressful life events. The tractography pipeline for the ABCD dataset has been described previously. The average FA for the left and right uncinate were estimated using procedures described previously, and then compared to the participant's life stressors behavioral measures by fitting a linear regression to the data.

**Structural networks.** Following tract segmentation, structural networks were generated using the multimodal 180 cortical node atlas and the tractograms for each participant using MRTrix3's tck2connectome (ref. 57) functionality implemented as A395 (<https://doi.org/10.25663/brainlife.app.395>). Connectomes were generated by computing the number of streamlines intersecting each ROI pairing in the 180 cortical node parcellation. Multiple adjacency matrices were generated, including count, density (that is, the count divided by the node volume of the ROI pairs), length, length density (that is length divided by the volume of the ROI pairs) and average and average density axial diffusivity, fractional anisotropy, mean diffusivity, radial diffusivity, neurite density index, orientation dispersion index and isotropic volume fraction. Density matrices were generated using the -invnodevol option<sup>58</sup>. For non-count measures (length, axial diffusivity, fractional anisotropy, mean diffusivity, radial diffusivity, neurite density index, orientation dispersion index, isotropic volume fraction), the average measure across all streamlines connecting and ROI pair was computed using MRTrix3's tck2scale functionality using the -precise option<sup>59</sup> and the -scale\_file option in tck2connectome. These matrices can be thought of as the 'average measure' adjacency matrices. These files were output as the 'raw' datatype and were converted to a conmat datatype using A393 (<https://doi.org/10.25663/brainlife.app.393>). Connectivity matrices were then converted into the 'network' datatype using functionality from Python functionality implemented as A335 (<https://doi.org/10.25663/brainlife.app.335>).

**Cortical and subcortical diffusion and morphometry mapping.** For the PING, HCP<sub>TR</sub>, HCP<sub>s1200</sub>, Cam-CAN and Indiana University Acute Concussion datasets, DTI and NODDI (if available) measures were mapped to each participant's cortical white matter parcels following methods found in Fukutomi and colleagues<sup>48</sup> using functions provided by Connectome Workbench<sup>60</sup> implemented as brainlife.io app A379 (<https://doi.org/10.25663/brainlife.app.379>). A Gaussian smoothing kernel (full-width at half-maximum ~4 mm,  $\sigma = 5/3$  mm) was applied along the axis normal to the midthickness surface, and DTI and NODDI measures were mapped using the wb\_command -volume-to-surface-mapping function. Freesurfer was used to map the average DTI and NODDI measures within each parcel using functionality from Connectome Workbench using A389 (<https://doi.org/10.25663/brainlife.app.389>) and A483 (<https://doi.org/10.25663/brainlife.app.483>). Measures of volume, surface area and cortical thickness for each cortical parcel were computed using Freesurfer and A464 (<https://doi.org/10.25663/brainlife.app.464>). Freesurfer was also used to generate parcel-average DTI and NODDI measures for the subcortical segmentation (aseg) from Freesurfer using A383 (<https://doi.org/10.25663/brainlife.app.383>). Measures of volume for each subcortical parcel were computed using Freesurfer and A272 (<https://doi.org/10.25663/brainlife.app.272>).

#### rs-fMRI preprocessing and functional connectivity matrix generation

For the HCP<sub>TR</sub> and Cam-CAN datasets, unprocessed resting-state functional MRI (rs-fMRI) datasets were preprocessed using fMRIPrep implemented as A160 (<https://doi.org/10.25663/brainlife.app.160>). Briefly, fMRIPrep does the following preprocessing steps. First, individual images are aligned to a reference image for motion estimation and correction using mcflirt from FSL. Next, slice timing correction is performed in which all slices are realigned in time to the middle of each relaxation time using 3dTShift from AFNI. Spatial distortions are then corrected using field map estimations. Finally, the fMRI data is aligned to the structural T1w image for each participant. Default parameters provided by fMRIPrep were used. For a subset of analyses involving the HCP test and retest datasets, the preprocessed rs-fMRI datasets provided by the HCP consortium were used. Following preprocessing via fMRIPrep for the volume data, connectivity matrices were generated using the Yeo17 parcellation and A369



(<https://doi.org/10.25663/brainlife.app.369>) and A532 (<https://doi.org/10.25663/brainlife.app.532>). Within-network functional connectivity for the 17 canonical resting-state networks was computed by computing the average functional connectivity values within all of the nodes belonging to a single network. These estimates were used for subsequent analyses.

### rs-fMRI gradient processing

For the HCP<sub>TR</sub> and Cam-CAN datasets, unprocessed rs-fMRI data from the HCP Test and Cam-CAN datasets were preprocessed using fMRIPrep implemented as A160 (<https://doi.org/10.25663/brainlife.app.160>). Within this app, the same preprocessing steps are undertaken as in A160 (<https://doi.org/10.25663/brainlife.app.160>), except for an additional volume-to-surface mapping using `mri_vol2surf` from FreeSurfer. The surface-based outputs were then used to compute gradients following methodologies outlined in ref. 61 for each participant in the HCP<sub>s1200</sub>, HCP<sub>TR</sub> and Cam-CAN datasets using A574 (<https://doi.org/10.25663/brainlife.app.574>) using diffusion embedding<sup>62</sup> and functions provided by BrainSpace<sup>63</sup>. More specifically, connectivity matrices were computed from surface vertex values within each node of the Schaffer 1,000 parcellation<sup>64</sup>. Cosine similarity was then computed to create an affinity matrix to capture inter-area similarity. Dimensionality reduction is then used to identify the primary gradients. A normalized-angle kernel was used to create the affinity matrix, from which two primary components were identified. Gradients were then aligned across all participants using a Procrustes alignment and joined embedding procedure<sup>61</sup>. Values from the primary gradient and the cosine distance used to generate the affinity matrices were used for subsequent analyses.

### MEG processing

For some analyses, raw resting-state-MEG time series data from the Cam-CAN dataset was filtered using a Maxwell filter implemented as A476 (<https://doi.org/10.25663/brainlife.app.476>) and median split using A529 (<https://doi.org/10.25663/brainlife.app.529>). For the remainder of the analyses, filtered data provided by the Cam-CAN dataset was used. For all MEG data, power-spectrum density profiles (PSD) were estimated using functionality provided by MNE-Python<sup>28,65</sup> implemented as A530 (<https://doi.org/10.25663/brainlife.app.530>). Following PSD estimation, peak alpha frequency was estimated using A531 (<https://doi.org/10.25663/brainlife.app.531>). Finally, PSD profiles were averaged across all nodes within each of the canonical lobes (frontal, parietal, occipital, temporal) using A599 (<https://doi.org/10.25663/brainlife.app.599>). Measures of PSD and peak alpha frequency were used for all subsequent analyses.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

All data derived and described in this paper are made available via the [brainlife.io](https://brainlife.io) platform as ‘Publications’. User data agreements are required for some projects, like data from the HCP, Cam-CAN, PING, ABCD and HBN datasets. The Indiana University Acute Concussion and Oxford University Choroideremia & Stargardt’s Disease Datasets are part of ongoing research projects and will be made available at a later stage. All other datasets are made freely available via the [brainlife.io](https://brainlife.io) platform. See Supplementary Table 6 for the [brainlife.io](https://brainlife.io)/pubs.

### Code availability

As part of the article, we are describing a total of nine platform components. All components are made publicly available and open source under MIT License. All the software for the platform components is listed in Supplementary Table 1. In addition, we share the code used

for the statistical analyses as Jupyter Notebooks (Supplementary Table 2). Finally, the Apps used and tested in this article are listed in Supplementary Table 3.

### References

23. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
24. Caron, B. et al. Collegiate athlete brain data for white matter mapping and network neuroscience. *Sci. Data* **8**, 56 (2021).
25. Yushkevich, P. A. et al. Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: towards a harmonized segmentation protocol. *Neuroimage* **111**, 526–541 (2015).
26. Karcher, N. R. & Barch, D. M. The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology* **46**, 131–142 (2021).
27. Yushkevich, P. A. et al. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum. Brain Mapp.* **36**, 258–287 (2015).
28. Tournier, J.-D. et al. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* **202**, 116137 (2019).
29. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
30. Benson, N. C. et al. The retinotopic organization of striate cortex is well predicted by surface topology. *Curr. Biol.* **22**, 2081–2085 (2012).
31. Benson, N. C., Butt, O. H., Brainard, D. H. & Aguirre, G. K. Correction of distortion in flattened representations of the cortical surface allows prediction of V1-V3 functional organization from anatomy. *PLoS Comput. Biol.* **10**, e1003538 (2014).
32. Cox, R. W. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* **29**, 162–173 (1996).
33. Ades-Aron, B. et al. Evaluation of the accuracy and precision of the diffusion parameter ESTimation with Gibbs and Noise removal pipeline. *Neuroimage* **183**, 532–543 (2018).
34. Andersson, J. L. R., Skare, S. & Ashburner, J. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage* **20**, 870–888 (2003).
35. Smith, S. M. et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, S208–S219 (2004).
36. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
37. Andersson, J. L. R., Graham, M. S., Zsoldos, E. & Sotiropoulos, S. N. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage* **141**, 556–572 (2016).
38. Andersson, J. L. R., Graham, M. S., Drobniak, I., Zhang, H. & Campbell, J. Susceptibility-induced distortion that varies due to motion: Correction in diffusion MR without acquiring additional data. *Neuroimage* **171**, 277–295 (2018).
39. Andersson, J. L. R. et al. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *Neuroimage* **152**, 450–466 (2017).
40. Jeurissen, B., Leemans, A. & Sijbers, J. Automated correction of improperly rotated diffusion gradient orientations in diffusion weighted MRI. *Med. Image Anal.* **18**, 953–962 (2014).
41. Tustison, N. J. et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *Neuroimage* **99**, 166–179 (2014).
42. Veraart, J. et al. Denoising of diffusion MRI using random matrix theory. *Neuroimage* **142**, 394–406 (2016).

43. Jenkinson, M. & Smith, S. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* **5**, 143–156 (2001).
44. Jenkinson, M., Bannister, P., Brady, M. & Smith, S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* **17**, 825–841 (2002).
45. Greve, D. N. & Fischl, B. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage* **48**, 63–72 (2009).
46. Pierpaoli, C., Jezzard, P., Basser, P. J., Barnett, A. & Di Chiro, G. Diffusion tensor MR imaging of the human brain. *Radiology* **201**, 637–648 (1996).
47. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000–1016 (2012).
48. Daducci, A. et al. Accelerated microstructure imaging via convex optimization (AMICO) from diffusion MRI data. *Neuroimage* **105**, 32–44 (2015).
49. Tournier, J.-D., Calamante, F. & Connelly, A. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. *Neuroimage* **35**, 1459–1472 (2007).
50. Yeh, F.-C., Wedeen, V. J. & Tseng, W.-Y. I. Generalized q-sampling imaging. *IEEE Trans. Med. Imaging* **29**, 1626–1635 (2010).
51. Yeh, F.-C. Shape analysis of the human association pathways. *Neuroimage* **223**, 117329 (2020).
52. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *Neuroimage* **62**, 1924–1938 (2012).
53. Bullock, D. et al. Associative white matter connecting the dorsal and ventral posterior human cortex. *Brain Struct. Funct.* <https://doi.org/10.1007/s00429-019-01907-8> (2019).
54. Sherbondy, A. J., Dougherty, R. F., Ben-Shachar, M., Napel, S. & Wandell, B. A. ConTrack: finding the most likely pathways between brain regions using diffusion tractography. *J. Vis.* **8**, 15.1–16 (2008).
55. Yoshimine, S. et al. Age-related macular degeneration affects the optic radiation white matter projecting to locations of retinal damage. *Brain Struct. Funct.* **223**, 3889–3900 (2018).
56. Yeh, F.-C. et al. Population-averaged atlas of the macroscale human structural connectome and its network topology. *Neuroimage* **178**, 57–68 (2018).
57. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. The effects of SIFT on the reproducibility and biological accuracy of the structural connectome. *Neuroimage* **104**, 253–265 (2015).
58. Hagmann, P. et al. Mapping the structural core of human cerebral cortex. *PLoS Biol.* **6**, e159 (2008).
59. Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. SIFT: Spherical-deconvolution informed filtering of tractograms. *Neuroimage* **67**, 298–312 (2013).
60. Van Essen, D. C. et al. The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
61. Margulies, D. S. et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl Acad. Sci. USA* **113**, 12574–12579 (2016).
62. Coifman, R. R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl Acad. Sci. USA* **102**, 7426–7431 (2005).
63. Vos de Wael, R. et al. BrainSpace: a toolbox for the analysis of macroscale gradients in neuroimaging and connectomics datasets. *Commun. Biol.* **3**, 103 (2020).
64. Schaefer, A. et al. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
65. Gramfort, A. et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).

## Acknowledgements

The brainlife.io project development and operations were supported by awards to F.P.: grant nos. NIH NIBIB R01EB029272, R01EB030896NSF and R01EB030896; NSF BCS 1734853 and 1636893; ACI 1916518, IIS 1912270; a gift from the Kavli Foundation; Wellcome Trust grant no. 226486/Z/22/Z and a Microsoft Investigator Fellowship. Additional funding was provided to support data collection used by the team, research that used brainlife.io or infrastructure that supported the platform: grant no. NIMH UM1NS132207 BRAIN CONNECTS: Center for Mesoscale Connectomics (Principal Investigator K. Ugurbil), grant no. NIMH R01MH133701 (C.R.). NSF grant award nos. 2004877 (S.V.-B.), 1541335 and 2232628 (S.M.), 1445604 and 2005506 (D.Y.H.), 1341698 and 1928224 (M. Norman), 1445606 (S.T.B.), 1928147 (S. Sanalevici). NIH grant award nos. 1U54MH091657 (HCP data, Principal Investigators D. Van Essen and K. Ugurbil), U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147 (ABCD Study, multiple Principal Investigators), P41EB017183 (J.V.), NIH NIBIB R01EB030896 (A.P.) and ANR-20-NEUC-0004-01 (multiple Principal Investigators). Multiple philanthropic contributions to the HBN (M. Milham).

## Author contributions

S.H. implemented most of the initial brainlife.io services. B.C. wrote the data analysis code, performed large-scale experiments, and prepared the figures and associated text. A.S.H. improved and implemented some of the services. J.F., R.C.C., D.H., D.S., D.P., L.K., J.K.L., C.R., F.N.-S., H.W., J.K.J., T.Z., J.W.K., S.K., C.V., D.N.B., B.M., D.B.A., F.D., J.G. and S.H., provided assets. All authors edited the manuscript. F.P. invented, designed and directed brainlife.io, wrote the paper and designed all the experiments and figures.

## Competing interests

F.P. received a Microsoft Faculty Fellowship, and Microsoft Azure sells Cloud Services. S.T.B. works for Hewlett-Packard Enterprise, which sells computing services. A.D.B. is an employee of BioSerenity, a company that develops medical devices to help diagnose and monitor patients with chronic diseases. S.H. is an employee of SHEGEL SPRL/BVBA a legal firm with expertise in data protection law. The other authors declare no competing interests.

## Additional information

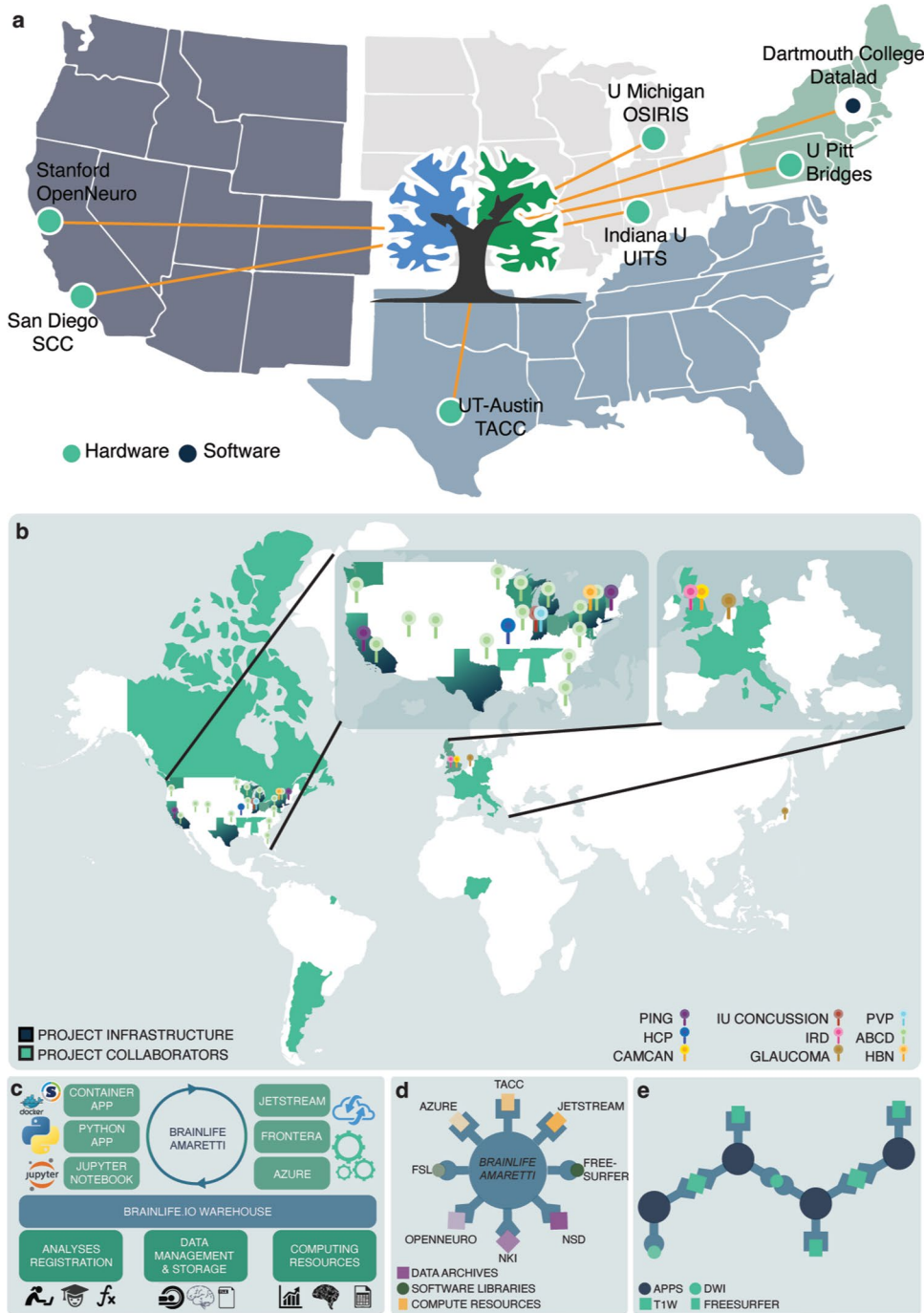
**Extended data** is available for this paper at <https://doi.org/10.1038/s41592-024-02237-2>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41592-024-02237-2>.

**Correspondence and requests for materials** should be addressed to Franco Pestilli.

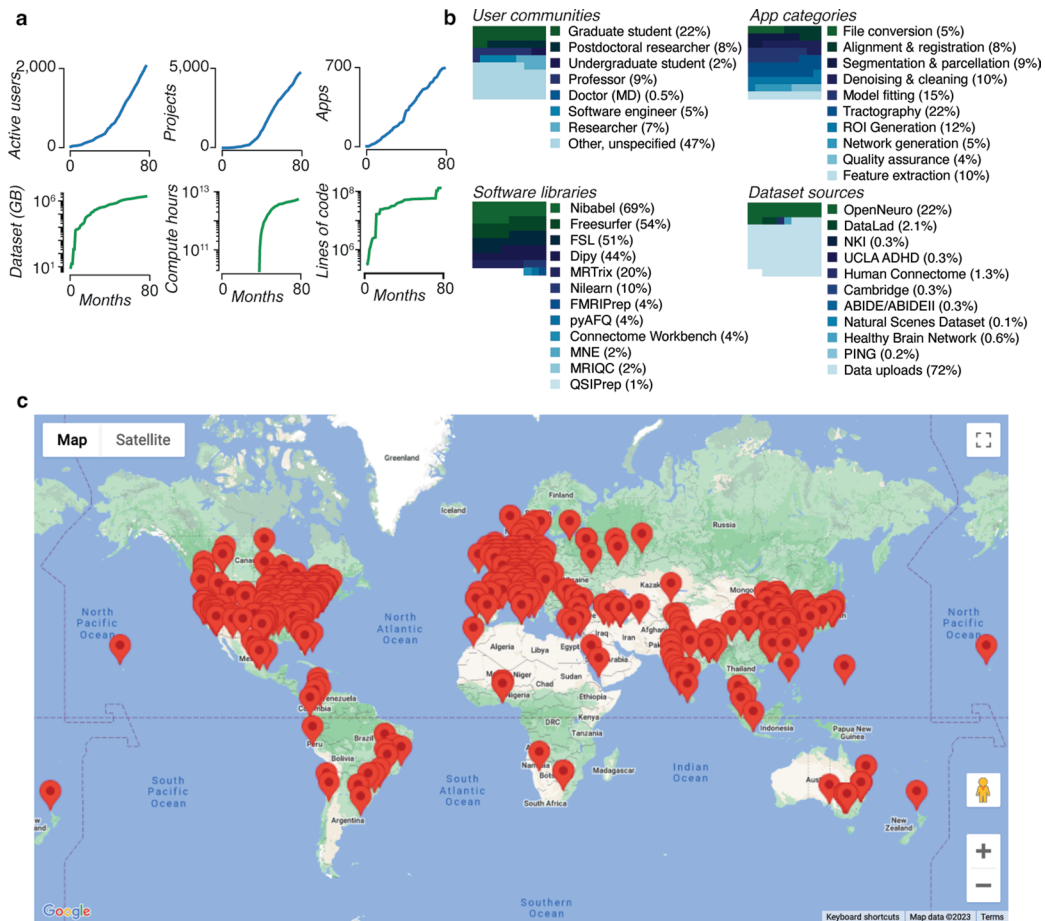
**Peer review information** *Nature Methods* thanks Jochem Rieger, Lucina Uddin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Nina Vogt, in collaboration with the *Nature Methods* team.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



**Extended Data Fig. 1 | Platform Architecture.** **a.** Map of the locations of critical hubs for brainlife.io. **b.** Map the locations of critical facets of this research, including project infrastructure (that is compute resources), collaborators, and data sources. As the United States and Europe are home to many of the infrastructural resources, collaborators, and data sources, more details for these regions are provided (insets). **c.** brainlife.io’s Amaretti links data archives, software libraries, and computing resources. Specifically, ‘Apps’ (containerized

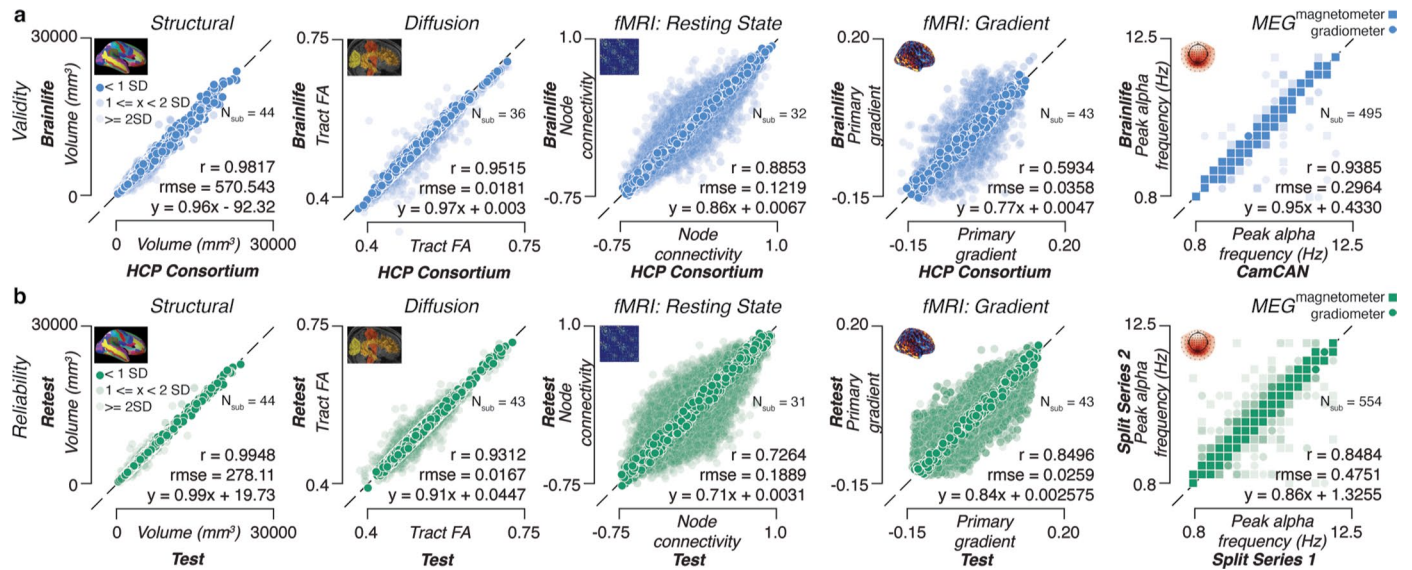
services defined on GitHub.com) are automatically matched with data stored in the ‘Warehouse’ with computing resources. Statistical analyses can be implemented using Jupyter Notebooks. **d.** brainlife.io provides efficient docking between data archives, processing apps, and compute resources via a centralized service. **e.** Apps use standardized Datatypes and allow ‘smart docking’ only with compatible data objects. App outputs can be docked by other Apps for further processing.



**Extended Data Fig. 2 | Platform Usage.** **a.** Top left. Number of users submitting more than 10 jobs per month. Top middle. Number of projects over time. Top right. Number of Apps over time. Bottom left. Data storage across all Projects. Bottom middle. Compute hours across all Projects (data only available 6 months post project start). Bottom right. Lines of code in the top 50 most-used Apps.

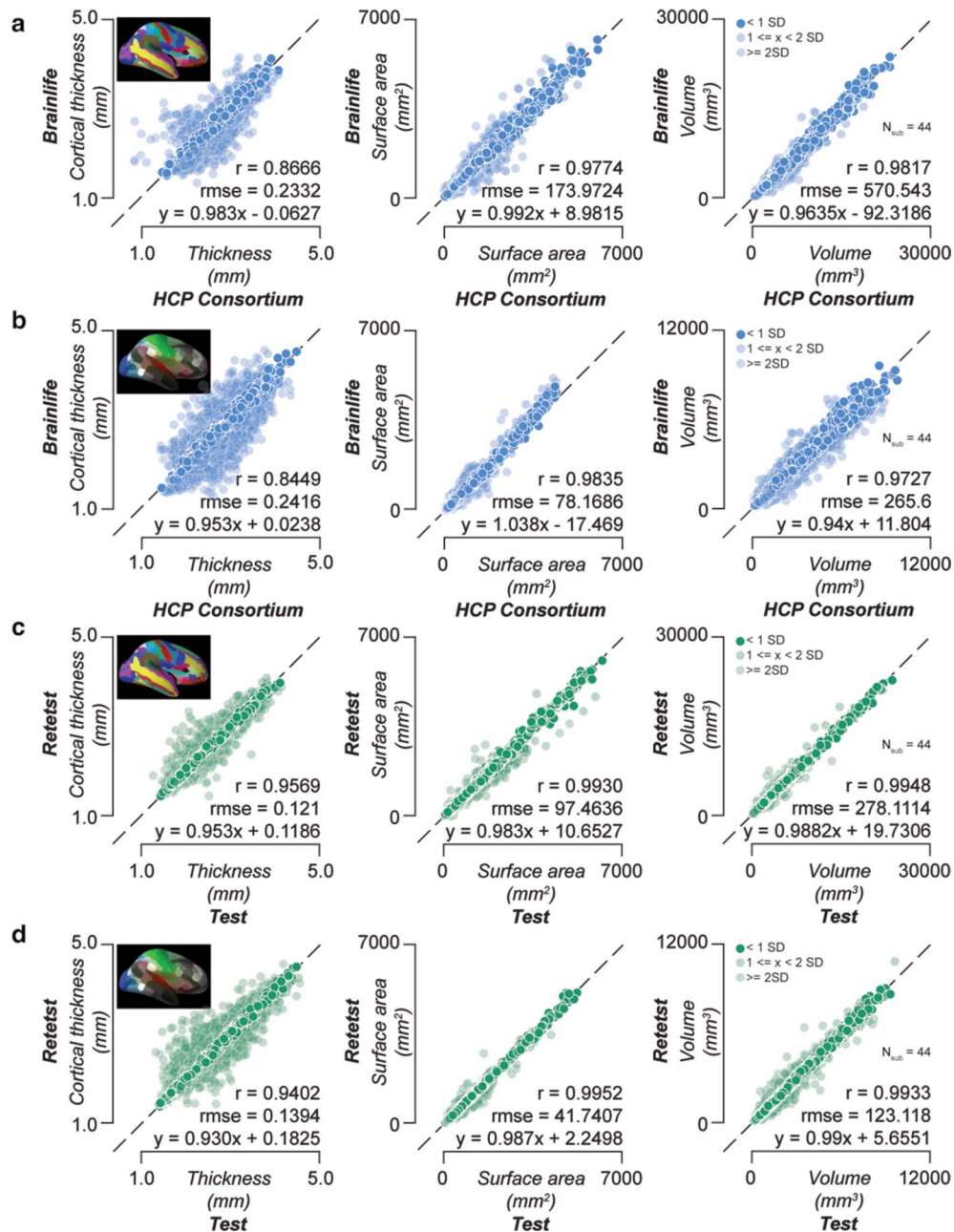
**b.** Top left. User communities. Top right. App categories. Bottom left. Percent of total jobs launched with the software library installed (percentage for jobs of top 50 most-used Apps). Bottom right. Datasets sources. **c.** Map of the locations of the users that created an account and accessed brainlife.io. This map is a proxy to the level of attention the platform achieved worldwide.





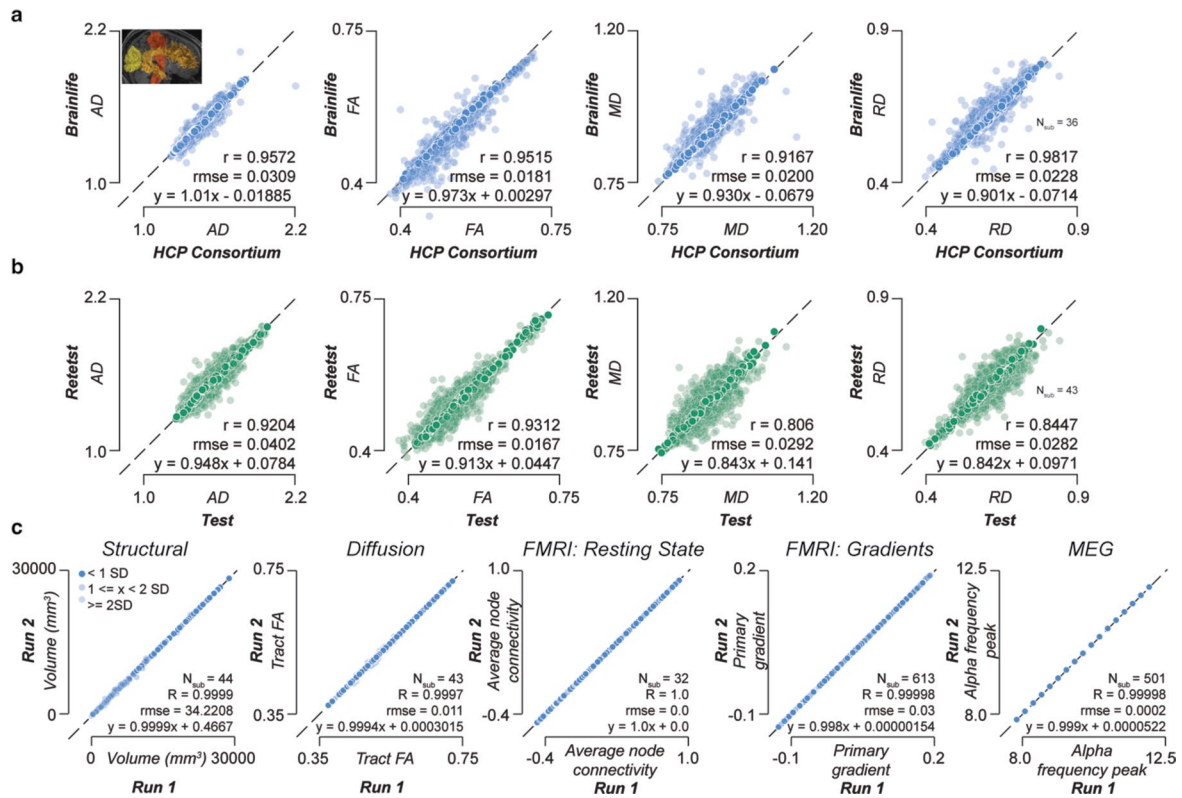
**Extended Data Fig. 3 | Data processing validity and reliability analysis.** Top row (a): Validity measures derived using the HCP Test-Retest (HCP<sub>TR</sub>) data. Each dot corresponds to the ratio for a given subject between data preprocessed and provided by the HCP Consortium vs data preprocessed on brainlife.io in a given measure for a given structure. Pearson's correlation (*r*), root mean squared error (*rmse*), and a linear fit between the test and retest results were calculated. Parcel volume (mm<sup>3</sup>). Tract-average fractional anisotropy (FA). Node-wise functional connectivity (FC)\*. Primary gradient value derived from resting-state fMRI\*. Peak frequency (Hz) in the alpha band derived from MEG. Data from magnetometer sensors are represented as squares, and data from gradiometer sensors are represented as circles. Dark colors represent data within ±1 standard deviation (SD). 50% opacity represents data within 1-2 SD. 25% opacity represents data outside 2 SD. \*A representative 5% of data presented. Bottom row (b): Test-retest

reliability measures derived from derivatives of the HCP<sub>TR</sub> dataset generated using brainlife.io. Each dot corresponds to the ratio between a test-retest subject and a given measure for a given structure. Pearson's correlation (*r*), root mean squared error (*rmse*), and a linear fit between the test and retest results were calculated. Parcel volume (mm<sup>3</sup>). Tract-average fractional anisotropy (FA). Node-wise functional connectivity (FC)\*. Primary gradient value derived from resting-state fMRI\*. Peak frequency (Hz) in the alpha band derived from MEG using the Cambridge (Cam-CAN) dataset. Data from magnetometer sensors are represented as squares, and data from gradiometer sensors are represented as circles. Dark colors represent data within ±1 standard deviation (SD). 50% opacity represents data within 1-2 SD. 25% opacity represents data outside 2 SD. \*A representative 5% of data presented.



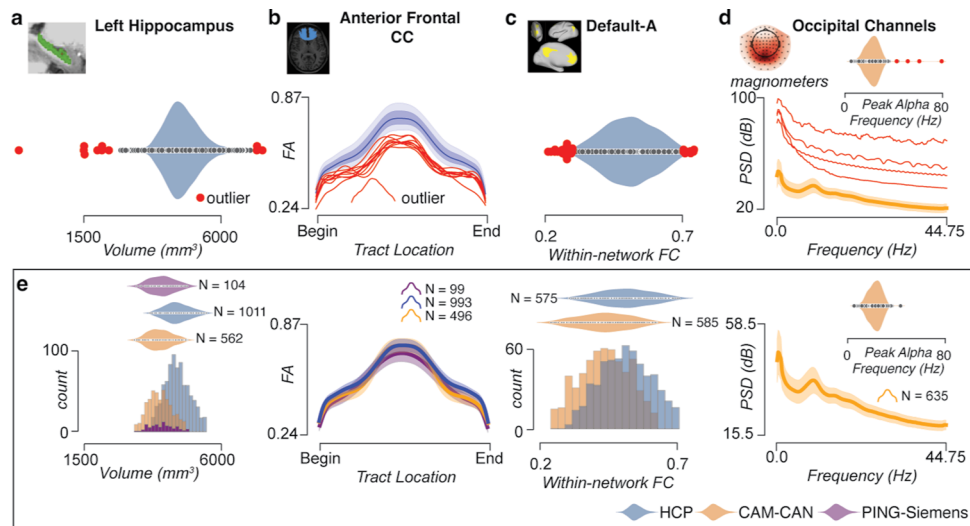
**Extended Data Fig. 4 | Processing with brainlife.io is valid and test-retest reliability is high - Structural MRI.** Top rows: Validity measures derived using the HCP<sub>TR</sub> data preprocessed and provided by the HCP Consortium compared to data preprocessed on brainlife.io. Each dot corresponds to the ratio for a given subject between data preprocessed and provided by the HCP Consortium vs data preprocessed on brainlife.io in a given measure for a given structure. Pearson's correlation ( $r$ ), root mean squared error ( $rmse$ ), and a linear fit between the test and retest results were calculated and provided. **a.** Destrieux Parcel thickness (mm), surface area (mm<sup>2</sup>), and volume (mm<sup>3</sup>). **b.** HCP-mmp Parcel thickness (mm), surface area (mm<sup>2</sup>), and volume (mm<sup>3</sup>). Dark colors represent data within  $\pm 1$  standard deviation. 50% opacity represents data within 1-2 standard deviations.

25% opacity represents data outside 2 standard deviations. Bottom rows: Test-retest reliability measures derived from derivatives of the HCPTR dataset generated using brainlife.io. Each dot corresponds to the ratio between a test-retest subject and a given measure for a given structure. Pearson's correlation ( $r$ ), root mean squared error ( $rmse$ ), and a linear fit between the test and retest results were calculated and provided. **c.** Destrieux Parcel thickness (mm), surface area (mm<sup>2</sup>), and volume (mm<sup>3</sup>). **d.** HCP-mmp Parcel thickness (mm), surface area (mm<sup>2</sup>), and volume (mm<sup>3</sup>). Dark colors represent data within  $\pm 1$  standard deviation. 50% opacity represents data within 1-2 standard deviations. 25% opacity represents data outside 2 standard deviations.



**Extended Data Fig. 5 | Processing with brainlife.io is valid, reliable, and reproducible.** Top row: Validity measures derived using the HCP<sub>TR</sub> data preprocessed and provided by the HCP Consortium compared to data preprocessed on brainlife.io. Each dot corresponds to the ratio for a given subject between data preprocessed and provided by the HCP Consortium vs data preprocessed on brainlife.io in a given measure for a given structure. Pearson's correlation ( $r$ ), root mean squared error ( $rmse$ ), and a linear fit between the test and retest results were calculated and provided. v. Tract average AD, FA, MD, and RD. Dark colors represent data within  $\pm 1$  standard deviation. 50% opacity represents data within 1-2 standard deviations. 25% opacity represents data outside 2 standard deviations. Bottom row: Test-retest reliability measures derived from derivatives of the HCP<sub>TR</sub> dataset generated using brainlife.io. Each dot corresponds to the ratio between a test-retest subject and a given measure

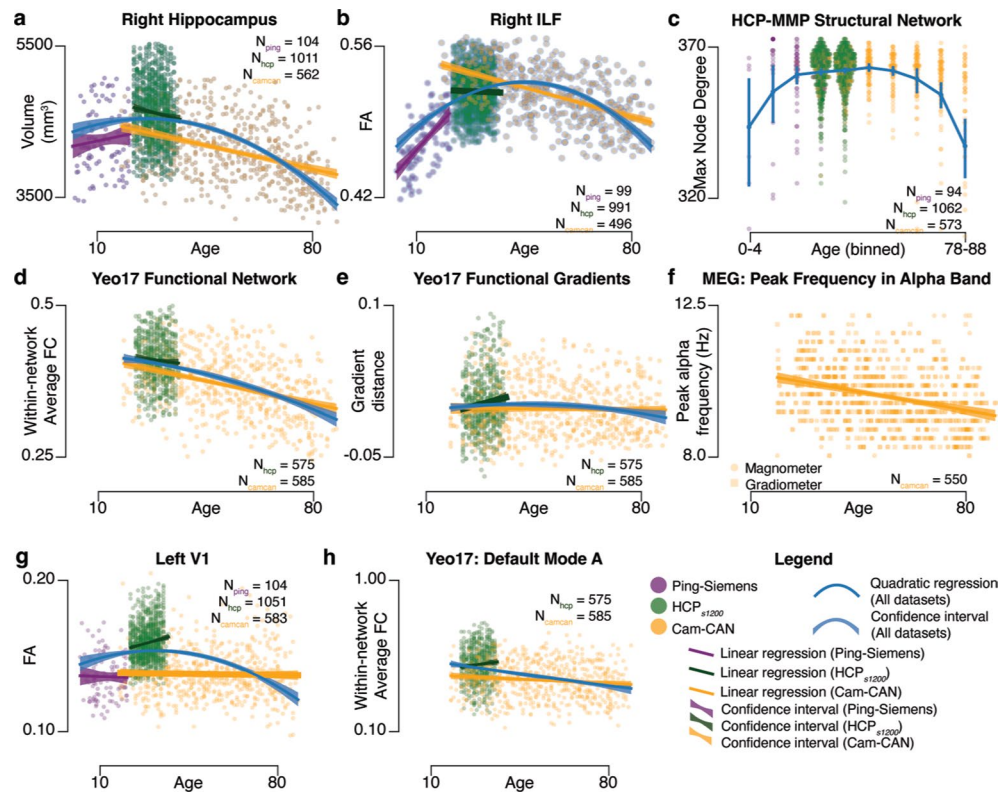
for a given structure. Pearson's correlation ( $r$ ), root mean squared error ( $rmse$ ), and a linear fit between the test and retest results were calculated and provided. w. Tract average AD, FA, MD, and RD. Dark colors represent data within  $\pm 1$  standard deviation. 50% opacity represents data within 1-2 standard deviations. 25% opacity represents data outside 2 standard deviations. c. Computational reproducibility values derived by repeating runs of brainlife.io Apps using the HCP<sub>TR</sub> dataset and the CAN dataset. Each dot corresponds to the ratio for a given subject between repeated runs of each App for a given structure. Pearson's correlation ( $r$ ), root mean squared error ( $rmse$ ), and a linear fit between the repeated runs was calculated. Destrieux Atlas Parcels volume (mm<sup>3</sup>). Tract-average fractional anisotropy (FA). Node-average functional connectivity (FC). Primary gradient values derived from resting state fMRI. Peak alpha frequency (Hz) in the alpha band derived from MEG.



**Extended Data Fig. 6 | Reference datasets for quality assurance.** Example workflow for building normative reference ranges for multiple derived statistical products (cortical parcel volume, white matter tract profilometry, within-network functional connectivity, and power-spectrum density (PSD)). **a.** Cortical volumes of the left hippocampus from HCP participants. Red dots indicate outlier data points. **b.** Average fractional anisotropy (FA) profiles (blue line) plotted with two standard deviations (shaded regions). Red lines indicate outlier profiles. **c.** Within-network functional connectivity for the nodes within the Default-A network using the Yeo17 atlas. Red dots indicate outlier data

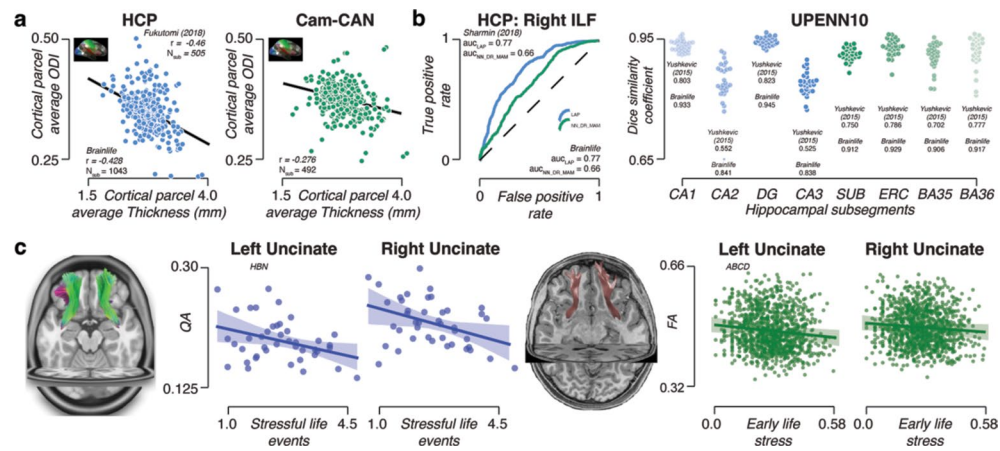
points. **d.** Average PSD from occipital channels using magnetometer sensors from Cam-CAN participants with one standard deviation (shaded regions). Red lines indicate outlier participants. Peak alpha frequency distribution was also computed, and outliers were detected (inset). **e.** Normative reference distributions for each derived statistical product across the PING (purple), HCP (blue), and Cam-CAN (orange) datasets. These distributions have had outliers removed. An example of the brainlife.io visualization for reference datasets can be found in Fig. S5. Data are presented as mean values  $\pm$  SEM.



**Extended Data Fig. 7 | Lifelong brain maturation estimated across datasets.**

Relationship between subject age and **a**. Right hippocampal volume, **b**. Right inferior longitudinal fasciculus (ILF) fractional anisotropy (FA), **c**. maximum node degree of density network derived using the hcp-mmp atlas, **d**\*. Within-network average functional connectivity (FC) derived using the Yeo17 atlas, and **e**\*. Functional gradient distance for visual resting state network derived from the Yeo17 atlas, and **f**. Peak frequency in the alpha band derived from magnetometer (squares) and gradiometers (circles) from MEG data. These analyses include subjects from

the PING (purple), HCPs1200 (green), and Cam-CAN (yellow) datasets. Linear regressions were fit to each dataset, and a quadratic regression was fit to the entire dataset (blue). \* All points in e, and f are presented. See Fig. 2a. Relationship between age of subject and **g**. Cortical fractional anisotropy (FA) of the left V1, **h**. Within-network average functional connectivity (FC) from the Yeo17 Default Mode - A network. These analyses include subjects from the PING (purple), HCPs1200 (green), and CAN (yellow) datasets. Linear regressions were fit to each dataset, and a quadratic regression was fit to the entire dataset (blue).



**Extended Data Fig. 8 | Replication of previous studies using brainlife.io.**

**a.** Average cortical hcp-mmp parcel thickness ( $N_{struc} = 322$ ) compared to parcel orientation dispersion index (ODI) from the NODDI model mapped to the cortical surface (inset) of the HCPS1200 dataset ( $N_{sub} = 1,043$ ) and Cam-CAN ( $N_{sub} = 492$ ) dataset compared to the parcel-average cortical thickness.

**b.** Receiver operator curves (ROC) comparing the performance of segmentation of the Right ILF using two automated segmentation methods (LAP: blue, NN\_DR\_MAM: green) in a subset of the HCPS1200 dataset ( $N_{sub} = 15$ ). Dice coefficients between manual and automated segmentation of the hippocampus using AHSS

method in UPENN dataset. **c.** Stressful life events obtained from Negative Life Events Schedule (NLES) survey from Healthy Brain Network participants ( $N_{sub} = 42$ ) compared to Uncinate-average normalized Quantitative Anisotropy (QA). Mean linear regression (blue line) fits and standard deviation (shaded blue). Early life stress was obtained from multiple surveys collected from ABCD participants ( $N_{sub} = 1,107$ ) compared to Uncinate-average Fractional Anisotropy (FA). Linear regression (green line) fits the data with standard deviation (shaded green). See Fig. 2b,c.

## Extended Data Table 1 | Platform microservices

Service	Description	GitHub Repos
UI	Platform endpoint, providing an user interface that integrates the diverse services in <a href="#">Brainlife</a>	<a href="https://github.com/brainlife/warehouse/tree/master/ui">https://github.com/brainlife/warehouse/tree/master/ui</a>
Warehouse	Data storage and management	<a href="https://github.com/brainlife/warehouse/">https://github.com/brainlife/warehouse/</a>
Amaretti	Automated scheduling service identifying appropriate compute resources and staging and archiving data	<a href="https://github.com/brainlife/amaretti/">https://github.com/brainlife/amaretti/</a>
ezBIDS	DICOM to BIDS conversion	<a href="https://github.com/brainlife/ezbids/">https://github.com/brainlife/ezbids/</a>
Vis	Services available for running visualizations within the platform	<a href="https://github.com/brainlife/brainlife/tree/master/vis">https://github.com/brainlife/brainlife/tree/master/vis</a>
Event	Event-driven integrator, to provide real-time feedback for users	<a href="https://github.com/brainlife/event/">https://github.com/brainlife/event/</a>
Service Monitoring	Monitors individual actions performed by the site	<a href="https://github.com/brainlife/service-monitor">https://github.com/brainlife/service-monitor</a>
CLI	Command-line interface for performing data manipulations and data scrubbing	<a href="https://github.com/brainlife/cli">https://github.com/brainlife/cli</a>
Auth	Centralized authentication for the multiple <a href="#">Brainlife</a> services	<a href="https://github.com/brainlife/auth">https://github.com/brainlife/auth</a>

Table with list of all platform services, name, scope, service URL (pointer to brainlife page if available as direct URL) and github URL for code.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection This article utilizes open data repositories all described in the methods section.

Data analysis No commercial software were used for data analysis. All data analyses were made using open source code, developed by the project or from other open source projects. Software packages used to process the data described in the manuscript include, but may not be limited to: Freesurfer, FSL, AFNI, ANTs, fMRIPrep, JSON, python, pandas, numpy, scipy, seaborn, QSIprep, Dipy, MRTrix3, Trekker, dsistudio, Nilearn, mne-python, connectome workbench, MatLab, Nibabel, pyAFQ, and MRIQC. All notebooks used to analyze the data can be found at [https://github.com/bp-notebooks/bl\\_paper](https://github.com/bp-notebooks/bl_paper). All code supporting the platform can be found at <https://github.com/brainlife>. Tables provided in Supplementary Information contain information about exact links to notebooks and platform code.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.



## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data derived and described in this paper are made available via the brainlife.io platform as "Publications". User data agreements are required for some projects, like data from the HCP, Cam-CAN, PING, ABCD, and HBN datasets. The Indiana University Acute Concussion Dataset and the Oxford University Choroideremia & Stargardt's Disease Dataset are parts of ongoing research projects and are not being released at this current time. All other datasets are made freely available via the brainlife.io platform. See Supplementary Table 6 for the brainlife.io/pubs [we have added one example data record (<https://doi.org/10.25663/brainlife.pub.40>) for the review process <the DOIs for the remaining data records will be added at publication>].

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	We used open repositories and other secondary data that have their own policies for reporting. We did not analyze subjects by either sex or gender.
Population characteristics	For some analyses, we analyzed the relationship between brain properties and age. For others, we analyzed the relationship between brain properties and clinical disorder diagnosis. These specific analyses can be found in Figure 1 (a, d) of the manuscript. For no other analyses did we include covariates.
Recruitment	We did not recruit subjects as we used secondary data.
Ethics oversight	All ethics review was performed at the sites of the original data collectors. We performed only secondary data analysis and were not involved in recruiting subjects.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	A multi-factorial study design was implemented using secondary data analyses. Some analyses involved using brain properties to test platform tool validity, reliability, and replicability. Other analyses involved using brain properties to identify those within specific clinical populations. All analyses were quantitative in scope.
Research sample	We used a large number of participants (~3000) in totum for all analyses from a variety of backgrounds (populations, scanner types). We used secondary data from many open projects, including the following: HCP, Cam-CAN, PING, ABCD, UPENN, and HBN. We have provided detailed descriptions of the data used from each cohort in the main manuscript online methods section.
Sampling strategy	We attempted to use all subjects available for each dataset. In certain situations, including the PING and ABCD datasets, only a subset of participants were used. For the PING, we only had access to the Siemen's scanner data at the time of analyses and manuscript preparation, leaving us with around 100 subjects. For the ABCD, we only used a random subset of ~1100 participants. Overall, because our results are not scientific in nature, these provided a large enough sample to demonstrate the platform's capabilities overall.
Data collection	We only performed secondary data analysis and did not collect new data for the study. Because of this, research blindness to experimental conditions or hypotheses were not required.
Timing	We only performed secondary data analysis and did not collect new data for the study.
Data exclusions	For the PING, we only had access to the Siemen's scanner data at the time of analyses and manuscript preparation, leaving us with around 100 subjects. For the ABCD, we only used a random subset of ~1100 participants. Overall, because our results are not

scientific in nature, these provided a large enough sample to demonstrate the platform's capabilities overall. Otherwise no data were excluded.

Non-participation

We only performed secondary data analysis and did not collect new data for the study. No participants were involved.

Randomization

We only performed secondary data analysis and did not collect new data for the study.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input type="checkbox"/>	<input checked="" type="checkbox"/> MRI-based neuroimaging

## Magnetic resonance imaging

### Experimental design

Design type

We only performed secondary data analysis and did not collect new data for the study.

Design specifications

We only performed secondary data analysis and did not collect new data for the study.

Behavioral performance measures

We only performed secondary data analysis and did not collect new data for the study.

### Acquisition

Imaging type(s)

functional, diffusion-weighted, anatomical, meg/eeg

Field strength

3 Tesla

Sequence & imaging parameters

Validity, reliability, reproducibility, replicability, developmental trends, & reference datasets  
 Human Connectome Project (HCP; Test-Retest, s1200-release) 23. Data from these projects were used to assess the validity, reliability, and reproducibility of the platform. They were used to assess the abilities of the platform to identify developmental trends in structural and functional measures, and they were used to generate reference datasets. Structural data (sMRI): The minimally-preprocessed structural T1w and T2w images from the Human Connectome Project (HCP) from 1066 participants from the s1200 and 44 participants from the Test-Retest releases were used. Specifically, the 1.25 mm 'acpc\_dc\_restored' images generated from the Siemens 3T MRI scanner were used for all analyses involving the HCP. For most examinations, the already-processed Freesurfer output from HCP was used. Diffusion data (dMRI): To assess the validity of preprocessing on brainlife.io, the unprocessed dMRI data from 44 participants from the HCP Test dataset was used. For reliability and all remaining analyses, the minimally-preprocessed diffusion (dMRI) images from 1,066 participants from the s1200 and 44 participants from the Test-Retest releases from the 3T Siemens scanner were used. All processes incorporated the multi-shell acquisition data. Functional data (fMRI): For validation, the unprocessed resting-state functional MRI (fMRI) from 44 participants from the HCP Test dataset was compared to the minimally-preprocessed BOLD data provided by HCP. For reliability and all other analyses, the minimally-preprocessed BOLD data from 1,066 participants from the s1200 and 44 participants from the Test-Retest releases from the 3T Siemens scanner were used.  
 The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) 7. The data from this project were used to assess the validity, reliability, and reproducibility of the platform and to assess the abilities of the platform to identify developmental trends of structural and functional measures, and to generate reference datasets. Structural data (sMRI): The unprocessed 1mm isotropic structural T1w and T2w images from 652 participants from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study were used. Diffusion data (dMRI): The unprocessed 2mm isotropic diffusion (dMRI) images from 652 participants from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study were used. Functional data (fMRI): The 3mm x 3mm x 4mm unprocessed resting-state fMRI images from 652 participants from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study were used. Electromagnetic data (MEG): The 1000 Hz resting-state filtered and unfiltered datasets from 652 participants from the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study were used.  
 Developmental trends & reference datasets  
 Pediatric Imaging, Neurocognition, and Genetics (PING) 11. The data from this project were used to assess the abilities of the platform to identify developmental trends of structural measures and to generate reference datasets. Structural

data (sMRI): The unprocessed 1.2 x 1.0 x 1.0 mm structural T1w and the 1.0 mm isotropic T2w images from 110 participants from the Pediatric Imaging, Neurocognition, and Genetics (PING) study were used. Diffusion data (dMRI): The unprocessed 2mm isotropic diffusion (dMRI) images from 110 participants from the Pediatric Imaging, Neurocognition, and Genetics (PING) study were used.

#### Replicability datasets

Adolescent Brain Cognitive Development (ABCD) 8,28. Structural data (sMRI): The unprocessed 1mm isotropic structural T1w and T2w images from a subset of 1,877 participants from the Adolescent Brain Cognitive Development (ABCD release-2.0.0) study were used. Diffusion data (dMRI): The unprocessed 1.77mm isotropic diffusion (dMRI) images from a subset of 1877 participants from the Adolescent Brain Cognitive Development (ABCD release-2.0.0) study were used. A single diffusion gradient shell was used for these experiments ( $b=3000s/ms^2$ ). Research approved by the University of Arkansas IRB (#2209425822).

Healthy Brain Network (HBN) 10. The data from this project were used to assess the abilities of the platform to replicate previously published findings via the assessment of the relationship between microstructural measures mapped to segmented uncinata fasciculi and self-reported early life stressors. Research approved by the University of Pittsburgh IRB (#PRO17060350). Structural data (sMRI): The 0.8 mm isotropic structural T1w images from 42 participants from the Healthy Brain Network (HBN) study were used. Diffusion data (dMRI): The unprocessed 1.8 mm isotropic diffusion (dMRI) images from 42 participants from the CitiGroup Cornell Brain Imaging Center site of the Healthy Brain Network (HBN) study were used. Research approved by the University of Pittsburgh IRB (#PRO17060350).

UPENN-PMC 29. The data from this project were used to assess the abilities of the platform to replicate previously published findings via the assessment of the performance of an automated hippocampal segmentation algorithm. All procedures were conducted under the approval of the Institutional Review Board at the University of Texas at Austin. Structural data (sMRI): The T1w and T2w data were provided within the Automated Segmentation of Hippocampal Subfields (ASHS) atlas29.

#### Clinical-identification datasets

Indiana University Acute Concussion Dataset. The data from this project were used to assess the abilities of the platform to identify clinical populations via the mapping of microstructural measures to the cortical surface. Neuroimaging was performed at the Indiana University Imaging Research Facility, housed within the Department of Psychological and Brain Sciences with a 3-Tesla Siemens Prisma whole-body MRI using a 64-channel head coil. Within this study, 9 concussed athletes and 20 healthy athletes were included. Research approved by Indiana University (IRB: 906000405). Structural data (sMRI): High-resolution T1-weighted structural volumes were acquired using an MPRAGE sequence: T1 = 900 ms, TE = 2.7 ms, TR = 1800 ms, flip angle = 9°, with 192 sagittal slices of 1.0 mm thickness, a field of view of 256 x 256 mm, and an isometric voxel size of 1.0 mm<sup>3</sup>. The total acquisition time was 4 minutes and 34 seconds. High-resolution T2-weighted structural volumes were also acquired: TE = 564 ms, TR = 3200 ms, flip angle = 120°, with 192 sagittal slices, a field of view of 240 x 256 mm, and an isometric voxel size of 1.0mm<sup>3</sup>. Total acquisition time was 4 minutes 30 seconds. Diffusion data (dMRI): Diffusion data were collected using single-shot spin-echo simultaneous multi-slice (SMS) EPI (transverse orientation, TE = 92.00 ms, TR = 3,820 ms, flip angle = 78 degrees, isotropic 1.5 mm<sup>3</sup> resolution; FOV = LR 228 mm x 228 mm x 144 mm; acquisition matrix MxP = 138 x 138. SMS acceleration factor = 4). This sequence was collected twice, one in the AP fold-over direction and the other in the PA fold-over direction, with the same diffusion gradient strengths and the number of diffusion directions: 30 diffusion directions at  $b = 1000 s/mm^2$ , 60 diffusion directions at  $b = 1,750 s/mm^2$ , 90 diffusion directions at  $b = 2,500 s/mm^2$ , and 19  $b = 0 s/mm^2$  volumes. The total acquisition time for both sets of dMRI sequences was 25 minutes and 58 seconds.

Oxford University Choroideremia & Stargardt's Disease Dataset. The data from this project was used to assess the abilities of the platform to identify clinical populations via mapping retinal-layer thickness via OCT and mapping of microstructural measures along optic radiation bundles segmented using visual field information (eccentricity). Neuroimaging was performed at the Wellcome Centre for Integrative Neuroimaging, Oxford with the Siemens 3T scanner. Research approved by the UK Health Regulatory Authority reference 17/LO/1540. Structural data (sMRI): High-resolution T1-weighted anatomical volumes were acquired using an MPRAGE sequence: T1 = 904 ms, TE = 3.97 ms, TR = 1900 ms, flip angle = 8°, with 192 sagittal slices of 1.0 mm thickness, a field of view of 174 mm x 192 mm x 192 mm, and an isometric voxel size of 1.0 mm<sup>3</sup>. The total acquisition time was 5 minutes and 31 seconds. Diffusion data (dMRI): Diffusion data were collected using EPI (transverse orientation, TE = 92.00ms, TR = 3600 ms, flip angle = 78 degrees, 2.019 x 2.019 x 2.0 mm<sup>3</sup> resolution; FOV = 210 mm x 220 mm x 158 mm; acquisition matrix MxP = 210 x 210, SMS acceleration factor = 3). This sequence was collected twice, one in the AP fold-over direction and the other in the PA fold-over direction. The PA fold-over scan contained 6 diffusion directions, 3 at  $b = 0 s/mm^2$  and 3 at  $b = 2000 s/mm^2$ , and was used primarily for susceptibility-weighted corrections. The AP fold-over scan contained 105 diffusion directions, 5 at  $b = 0 s/mm^2$ , 51 at  $b = 1000 s/mm^2$ , and 49 at  $b = 2000 s/mm^2$ . The total acquisition time for both sets of dMRI sequences was 7 minutes and 8 seconds.

Area of acquisition

Brain

Diffusion MRI



Used



Not used

Parameters

A variety of diffusion MRI sources were used and are described above.

## Preprocessing

Preprocessing software

We used a variety of processing software, including but not limited to: Freesurfer, FSL, AFNI, ANTs, fMRIprep, JSON, python, pandas, numpy, scipy, seaborn, QSIprep, Dipy, MRtrix3, Trekker, dsistudio, Nilearn, mne-python, connectome workbench, MatLab, Nibabel, pyAFQ, and MRIQC.

Normalization

Linear and non-linear regression were used.

Normalization template

Various templates were depending on the analysis. The FreeSurfer FS-average in MNI coordinates, the Yeo17 atlas and the Human Connectome Multimodal Parcellation

Noise and artifact removal

Volume censoring

### Statistical modeling & inference

Model type and settings

Effect(s) tested

Specify type of analysis:  Whole brain  ROI-based  Both

Anatomical location(s)

Statistic type for inference  
(See [Eklund et al. 2016](#))

Correction

### Models & analysis

n/a | Involved in the study

Functional and/or effective connectivity

Graph analysis

Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Graph analysis