DATA DESCRIPTOR

TITLE

Cellular Resolution Cortico-Cortical Connectome of the Marmoset Monkey

AUTHORS

Piotr Majka[1,2,3] , Shi Bai[2,3] , Sophia Bakola[2,3,4] , Sylwia Bednarek[1] , Jonathan M. Chan[2,3] , Natalia Jermakow[1] , Lauretta Passarelli[4] , David H. Reser[3,5] , Panagiota Theodoni[6,7,8] , Katrina H. Worthy[3] , Xiao-Jing Wang[6] , Daniel K. Wójcik[1], Partha P. Mitra[9], Marcello G.P. Rosa[2,3]

AFFILIATIONS

1. Laboratory of Neuroinformatics, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, 02-093 Warsaw, Poland
2. Australian Research Council, Centre of Excellence for Integrative Brain Function, Monash University Node, Clayton, VIC 3800, Australia
3. Biomedicine Discovery Institute and Department of Physiology, Monash University, Clayton, VIC 3800, Australia
4. Department of Biomedical and Neuromotor Sciences Pharmacy and Biotechnology, University of Bologna, Bologna 40126, Italy.
5. Graduate Entry Medicine Program, Monash Rural Health-Churchill, School of Rural Health, Monash University, Churchill, VIC 3842, Australia.
6. Center for Neural Science, New York University, NY 10003, USA.
7. New York University - East China Normal University Institute of Brain and Cognitive Science, Shanghai 200241, China.
8. New York University Shanghai, Shanghai 200122, China.
9. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, NY 11724, USA

CORRESPONDING AUTHOR(S): Piotr Majka: p.majka@nencki.edu.pl

SUMMARY

This dataset presents the results of 143 injections of fluorescent retrograde tracers in the 53 brain hemispheres of the common marmoset neocortex. Data obtained from different animals are registered in a common stereotaxic space. The dataset is provided in various forms and data formats, facilitating a broad range of analyses. This includes, for instance, connectivity patterns relative to cytoarchitectural areas, featuring statistical properties such as the fraction of labeled neurons and the percentage of supragranular neurons. It also provides purely spatial (parcellation-free) data based on the stereotaxic coordinates of almost 2 million labeled neurons.

VERSION SPECIFICATIONS:

This is the initial version of the dataset, released February 28, 2020, tagged v1.

MATERIALS AND METHODS

Surgical procedures

Release v1 of the Marmoset Brain Connectivity Atlas includes the results of 143 injections of retrograde tracers in 52 young adult (1.4–4.6 years, median age: 2.5 years) marmosets (31 male, 21 female). Detailed metadata for each animal is available in the dataset's metadata directory in JSON files. All experiments conformed to the Australian Code of Practice for the Care and Use of Animals for Scientific Purposes and were approved by the Monash University Animal Experimentation Ethics Committee. Intramuscular (i.m.) injections of atropine (0.2 $mg \cdot kg^{-1}$) and diazepam (2 $mg \cdot kg^{-1}$) were administered as pre-medication, before each animal was anaesthetized with alfaxalone (10 $mg \cdot kg^{-1}$, i.m.) 30 min later. Dexamethasone (0.3 $mg \cdot kg^{-1}$, i.m.) and amoxicillin (50 $mg \cdot kg^{-1}$, i.m.) were also administered before positioning the animals in a stereotaxic frame. Body temperature, heart rate, and blood oxygenation (pO2) were continually monitored during surgery, and when necessary, supplemental doses of anesthetic were administered to maintain areflexia. Small incisions of the dura mater were made over the intended injection sites.

Six types of fluorescent tracers were used: fluororuby (FR; dextran-conjugated tetramethylrhodamine, molecular weight 10,000, 15% in dH2O), fluoroemerald (FE; dextran-conjugated fluorescein, molecular weight 10,000, 15% in dH2O), fast blue (FB, 2% in dH2O), diamidino yellow (DY, 2% in dH2O), and cholera toxin subunit B (CTB, conjugated with either Alexa 488 [CTBgr] or Alexa 594 [CTBr], 1% in PBS). The tracers were injected using 25 µl constant rate microsyringes (Hamilton, Reno, NV) fitted with a fine glass micropipette tip. Each tracer was injected over 15–20 min, with small deposits of tracer made at different depths. Following the last deposit, the pipette was left in place for 3–5 min to minimize tracer reflux. After the injections, the surface of the brain was covered with moistened ophthalmic film, over which the dural flaps were carefully arranged. The excised bone fragment was repositioned and secured in place with dental acrylic, and the wound was closed in anatomical layers. Postoperative injectable analgesics were administered immediately after the animal exhibited spontaneous movements (Temgesic 0.01 $mg \cdot kg^{-1}$, i.m., and Carprofen 4 $mg \cdot kg^{-1}$, s.c.), followed by oral Metacam (0.05 $mg \cdot kg^{-1}$) for 3 consecutive days.

Histological processing

Survival times varied between 3 and 22 days (median: 15 days), after which the animals were anesthetized with alfaxalone (10 $mg \cdot ml^{-1}$ i.m.) and, following loss of consciousness, administered an overdose of sodium pentobarbitone (100 $mg \cdot kg^{-1}$, i.v.). They were then immediately perfused through the heart with 500 ml of heparinized saline, followed by 500 ml of 4% paraformaldehyde (PFA) in 0.1 M phosphate buffered saline (PBS; pH 7.4). The brains were post-fixed in the same medium for at least 24 h, and then immersed in buffered PFA with increasing concentrations of sucrose (10–30%). They were then sectioned (40 µm thickness) in the coronal (most cases) or parasagittal (three hemispheres) plane, using a cryostat. One section in five was mounted unstained for examination of fluorescent tracers, and coverslipped after quick dehydration (2 × 100% ethanol) and defatting (2 × xylene). Adjacent sections were stained for cell bodies (using the cresyl violet stain, or the NeuN stain), cytochrome oxidase, or myelin. Hence, the apparent spacing between adjacent sections in each series was 200 µm. Stained sections were scanned using an Aperio Scanscope AT Turbo system (Leica Biosystems), providing a resolution of 0.50 $\mu m \cdot pixel^{-1}$.

Microscopic analysis and digitalization
Sections were examined using epifluorescence microscopes. Labeled neurons were identified using ×10 or ×20 dry objectives, and their locations within the cortex and subcortical structures were mapped using a digitizing system attached to the microscope. Labeled cells were accepted as valid only if a nucleus could be discerned to minimize the problem of overestimating the number of neurons due to the inclusion of cytoplasmic fragments.

Reference brain atlas
We selected the Paxinos et al. (2012) stereotaxic atlas of the marmoset brain as the three-dimensional reference space. The system of coordinates in this atlas is based on cranial landmarks: the horizontal zero plane is defined as the plane passing through the lower margin of the orbits and the center of the external auditory meatuses, the anteroposterior zero plane is defined as the plane perpendicular to the horizontal zero plane which passes the centers of the external auditory meatuses, and the left–right zero plane is the midsagittal plane.
The Nissl-stained plates illustrated in the PDF edition of the book were converted into a 3D Nifti file (available in the `paxinos_et_al_2012_mbisc_atlas` directory of the dataset). This resulted in a volumetric image of a resolution of 40 × 500 × 40 µm (mediolateral, rostrocaudal, and dorsoventral, respectively), which preserves the stereotaxic coordinates of the source atlas plates. The atlas parcellation scheme consists of 116 cortical areas.

3D reconstruction and mapping procedure
A computational workflow was established to register the results of individual experiments into the reference template (Majka et al. 2016). The pipeline computes a set of spatial transformations that allow the location of cell bodies to be expressed within a common set of stereotaxic coordinates. Three main steps can be distinguished: The affine reconstruction step, the deformable reconstruction, and the combined, affine and deformable, co-registration with the template image (see Majka et al. 2016 for a detailed description). The pipeline is based on the Possum 3D reconstruction framework (Majka and Wójcik 2016; https://github.com/pmajka/poSSum) and the Advanced Normalization Tools (ANTS) software suite (Avants et al., 2011; RRID: SCR_004757; http://stnava.github.io/ANTs/)

Quantification of connectivity patterns
The locations of cells and injection sites were mapped into the template using the computed set of spatial transformations. Subsequently, they were assigned to a cortical area based on the stereotaxic location of the corresponding voxel(s) relative to the atlas parcellation.
Seventy-nine injection sites were entirely contained within estimates of a single cytoarchitectural area, 41 were >80% contained within an area, and 23 likely crossed borders. The assignment to an area was based on the voxel containing the injection barycenter and validated by an expert. The injections ' metadata contains detailed remarks regarding the injected area(s).

In addition to being assigned to an area, labeled cells were divided into either those above layer 4 (supragranular) or below it (infragranular) by a procedure that involved manual delineation of the granular cell layer across the entire set of sections that contained labeled neurons. Cells located in the entorhinal cortex (Ent), piriform cortex (Pir), amygdalopiriform transition area (APir), and area 29a-c (A29a-c) were excluded from this assignment due to the lack of a visible layer 4. In the intermediate and medial sectors of the primary motor cortex (A4ab) and dorsocaudal premotor area (A6DC), the interface between layers 3 and 5 was used. The 3D locations, assigned areas, and laminar positions were stored in a database (`marmoset_brain_connectivity_1_0_master_database.db` file of the dataset).

The connectivity patterns obtained for each injection were quantified based on these data. The strength of a directed connection to an injected area A from an area B was defined as the Fraction of Labeled neurons (extrinsic; FLNe):

$$\mathrm{FLNe}_{B \to A} = \frac{\text{Number of neurons projecting to area } A \text{ from area } B}{\text{Total number of neurons projecting to area } A \text{ from all areas } - \text{neurons identified in area } A}$$

The weights of the within-area connections (i.e., from area A to itself) were not considered. The fraction of neurons that originate in the supragranular layers (supragranular neurons, SLN) of the source area was also computed:

$$\mathrm{SLN}_{B \to A} = \frac{\text{Number of supragranular neurons projecting to area } A \text{ from area } B}{\text{Total number of neurons projecting to area } A \text{ from area } B}$$

The FLNe and SLN values were calculated for each injection separately. An average FLNe value was obtained using the arithmetic mean for areas where multiple injections were placed. A corresponding average SLN value was obtained by first summing supragranular and infragranular neurons across injections in which a specific projection was observed, then calculating the SLN off the obtained totals.

USAGE NOTES

First, we would like to indicate that the Marmoset Brain Connectivity Atlas is a rather complex dataset in which the connectivity patterns are expressed in different ways using various data formats. Therefore, we advise paying close attention to the descriptions of the individual components of the dataset to avoid confusion and errors.

Master Database

The dataset's main file is the Marmoset Brain Connectivity Atlas Master Database, provided in SQLite format (`marmoset_brain_connectivity_1_0_master_database.db`). The database can be explored using many GUI clients, such as DB Browser for SQLite (https://sqlitebrowser.org/) or SQLiteStudio (https://sqlitestudio.pl/). The purpose of the master database is primarily to store the unaggregated and aggregated results and metadata in a logically and formally organized manner. Due to its general form, the database should not be used as a go-to dataset for exploring the connectivity patterns, but rather to compile or aggregate the data in a novel way, which is not currently available within this dataset. Figure 1 presents the overall structure of the database. A detailed description of each table and field is beyond the scope of this data descriptor document and can be found on the Marmoset Brain Connectivity Atlas Github Wiki.

Reference Brain Template and Parcellation

The `paxinos_et_al_2012_mbisc_atlas` contains the Marmoset Brain in Stereotactic Coordinates (Paxinos et al., 2012) stereotaxic atlas of the marmoset brain in NIfTI format. All stereotaxic coordinates available in this dataset are defined in the space of this atlas. The `atlas.nii.gz` contains a 3-dimensional RGB image of the left hemisphere, the `atlas_segmentation.nii.gz` contains the segmentation, and the `atlas_mask.nii.gz` file contains the binary mask of the atlas. Finally, the `atlas_labels.txt` file contains the descriptions of the segmentation labels of 136 cortical areas available in the parcelation.

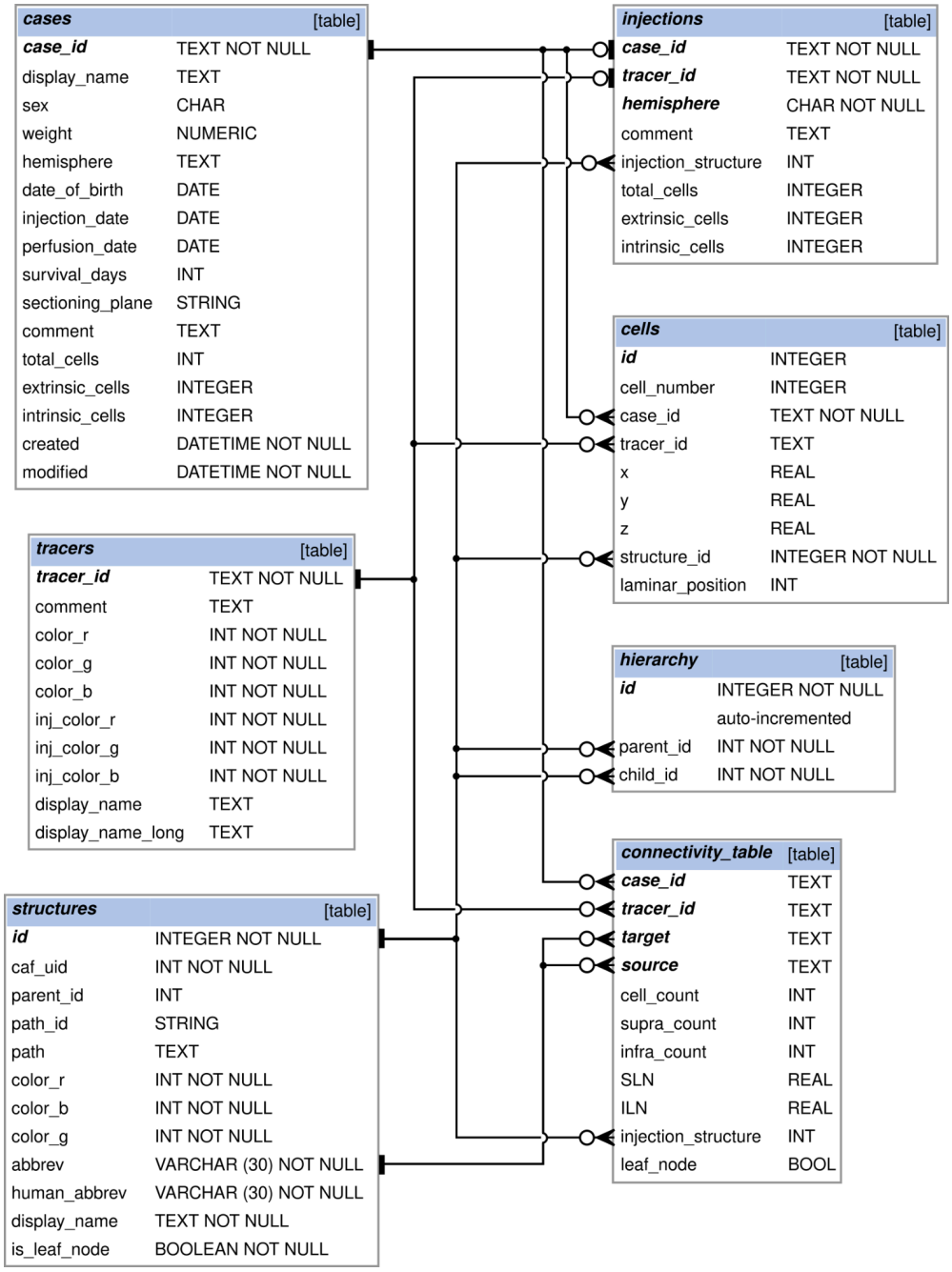

Figure 1. Overall structure of the Marmoset Brain Connectivity Atlas master database. A detailed description of each table and field is beyond the scope of this data descriptor document and can be found on the [Marmoset Brain Connectivity Atlas Github Wiki].

Aggregated Interareal Connectivity
Files in the `fine_per_area` directory contain data on aggregated connectivity for cortical areas with at least one injection. Results for areas that received multiple injections are averaged as explained earlier in this document. Individual CSV files have four columns with the following meaning: (1) target: area receiving the projection, (2) source: area sending the projection, (3) FLNe: fraction of extrinsic labelled neurons, (4) log10_flne: log10(FLNe). Lines starting with the hash (#) character are comments and contain no actual data.

Summarized results for individual injections
The `fine_per_injection` directory contains summarized (at the level of cortical areas) results for individual injections. Each CSV file contains results for a single injection (hence 143 files in total) and adheres to the following structure: (1) case_id: animal identifier, (2) tracer_id: tracer type, (3) target: area receieving the projection, (4) source: area sending the projection, (5) cell_count: the actual number of labelled cells identified in the source area, (6) FLNe: fraction of extrinsic labelled neurons, (7) log10_flne: log10(FLNe). Lines starting with the hash (#) character are comments and contain no actual data.

Cellular-resolution results for individual injections
The CSV files in the `individual_cells_per_injection` contain the cellular-resolution results for individual injections (i.e., each file provides a record for all labelled neurons identified in a given injection). The files adhere to the following structure: (1) case_id: animal identifier, (2) tracer_id: tracer type, (3) target: area receiving the projection, (4) source: area sending the projection, (5) mediolateral stereotaxic coordinate (mm), the values increase towards the lateral direction, (6) rostrocaudal stereotaxic coordinate (mm), positive values are caudal to the interaural line while negative values are rostral to the interaural line, (7) dorsoventral stereotaxic coordinate (mm), the values increase towards the dorsal direction from the interaural line (8) laminar position of the specific cell with respect to the granular cell layer: -1 for infragranular cells, +1 for supragranular cells, 0 for undefined or irrelevant.

Metadata for individual brain hemispheres
The `metadata` directory contains JSON files detailing individual brain hemispheres and all injections into a particular hemisphere. The files are organized into nested lists and provide the following metadata information: (1) animal identifier, (2) date of birth, (3) injected hemisphere: left or right, (4) list of all injections into the hemisphere (see below), (5) free comments related to surgeries and injections for better interpretation of the results, (6) perfusion date, (7) sectioning plane, (8) sex, (9) survival time, i.e. number of days between the surgery and the perfusion date, (10) weight in grams on the day of the surgery. Each element of the injection list contains the following information: (1-3) stereotaxic coordinates of the barycenter of the injection site following Paxinos et al. (2012) definition, (4) injected hemisphere, (5) free comments related to the injection, (6) abbreviation of the injected cortical area, (7) name of the Nissl-stained section with the location of the barycenter of the injection site, (8) identifier of the injected tracer.

Spatial distribution patterns for individual injections
The `nifti_maps_per_injection` directory contains spatial distribution patterns of labelled neurons for individual injections **after** mapping to the stereotaxic reference template. Each zip file contains three NIfTI files: one with all cells identified in a given injection, one for supragranular cells only, and one with the infragranular cells. NIfTI has a 200 µm per voxel isotropic resolution, and the value in each voxel correspond to the number of labelled neurons

at a particular location. Depending on the file, the sum of all voxels in a map corresponds to the total number of all supragranular or infragranular neurons. The maps are compatible (i.e., can be overlaid) on top of the NIfTI file with the reference template.

Data summaries

The dataset also provides summaries of all injections compiled in a few ways:

`marmoset_brain_connectivity_1_0_fln_matrix.txt`: Traditional interareal connectivity matrix. Each row of the file includes information on (1) source, (2) target, and (3) fraction of the extrinsic labeled neurons (FLNe) calculated for all injections into the given source area.
`marmoset_brain_connectivity_1_0_connectivity_matrix.gpickle`: Connectivity matrix in graph object in Python pickle format compatible with Python NetworkX package.

`marmoset_brain_connectivity_1_0_all_injections.txt`

A detailed tabular summary of all injections for all areas. The file has the following structure:

1) **`row_idx`**: running number for a record. The file has 16588 records (116 areas × 143 injections).
2) **`injection_idx`**: running number for an injection. The values go from 1 to 143.
3) **`case_id`**: identifier of an animal.
4) **`tracer_id`**: identifier of an injected tracer.
5) **`target`**: area receiving the projection (abbreviation).
6) **`source`**: area sending the projection (abbreviation).
7) **`cell_count`**: number of labelled neurons identified in the target area.
8) **`flne`**: fraction of extrinsic labelled neurons corresponding to the cell count.
9) **`gmeannz`**: geometric mean of non-zero FLNe values for areas that received multiple injections. In case there was only one injection, this value is equal to flne.
10) **`mean`**: Arithmetic mean for all FLNe values for areas that received multiple injections. In case there was only one injection, this value is equal to flne.
11) **`std`**: standard deviation for FLNe values for areas that received multiple injections. In case there was only one injection, the value is not provided.
12) **SLNe**: fraction of supragranular neurons within a particular connection.
13) **SLNeA**: Arithmetic mean for all SLNe values for areas that received multiple injections. In case there was only one injection, this value is equal to SLNe.

The file **`marmoset_brain_connectivity_FLN.json`** contains precisely the same data, but is provided in JSON format for more convenient programmatic processing.

DATA RECORDS

Directory structure:

/ **flne_per_area**
- / `flne_area_{area_abbreviation}.csv`
  Aggregated connectivity patterns for cortical areas with at least one injection.

/ **flne_per_injection**
- / `flne_injection_{case_id}-{tracer_id}.csv`
  Summarized results for individual injections.

/ **individual_cells_per_injection**
- / `indivdual_cells_injection_{case_id}-{tracer_id}.csv`
  Cellular-resolution results for individual injections.

/ **metadata**
- / `metadata_{case_id}.json`
  Metadata for individual brain hemispheres.

/ **nifti_maps_per_injection**
- / `nifti_map_{case_id}-{tracer_id}.zip`
  Spatial distribution patterns for individual injections and related injections.

/ **paxinos_et_al_2012_mbisc_atlas**
Marmoset Brain in Stereotactic Coordinates (Paxinos et al., 2012) atlas of the marmoset brain converted from PDF to NIfTI format.
- / `atlas.nii.gz`
  3-dimensional RGB image of the left hemisphere
- / `atlas_labels.txt`
  Descriptions of the segmentation into cortical areas.
- / `atlas_mask.nii.gz`
  Binary mask of the atlas
- / `atlas_segmentation.nii.gz`
  Segmentation into cortical areas

/ `marmoset_brain_connectivity_1_0_master_database.db`
Marmoset Brain Connectivity Atlas Master Database in SQLite format

/ `marmoset_brain_connectivity_1_0_all_injections.txt`
A detailed tabular summary of all injections for all areas. See “USAGE NOTES” for details.

/ `marmoset_brain_connectivity_FLN.json`
The summary is the same as above, but in JSON format.

/ `marmoset_brain_connectivity_1_0_fln_matrix.txt`
Traditional interareal connectivity matrix

/ `marmoset_brain_connectivity_1_0_connectivity_matrix.gpickle`
Connectivity matrix in Python pickle format compatible with the Python NetworkX package.

List all file formats and include which software created the respective files:

Format	Extension	Software used / file specification
Neuroimaging Informatics Technology Initiative NIfTI-1	.nii.gz	Recommended software: ITK-SNAP, 3D Slicer, Advanced Normalization Tools (ANTS) or Convert 3D for command line processing nibabel package for processing in Python Specification available at: https://nifti.nimh.nih.gov/nifti-1/index.html
Tab- or comma-separated values	.txt, .csv	Human-readable text file format that uses commas or tabulators to separate values and newlines to separate records. Suggested software: LibreOffice Calc or Python Pandas package for programmatic access

Graph object in Python pickle format	.gpickle	You can read and write NetworkX graphs as Python pickles:
		`import pickle`
		`with open('test.gpickle', 'rb') as f:`
		`...  G = pickle.load(f)`

SQLite database file format	.db	The database can be explored using many GUI clients, such as DB Browser for SQLite (https://sqlitebrowser.org/) or SQLiteStudio (https://sqlitestudio.pl/). For programmatic access, we recommend the Python sqlite3 package.
JavaScript Object Notation file format	.json	Human-readable text to store and transmit data objects. We recommend using a Python JSON encoder and decoder to manipulate the files.
ITK-SNAP label description file format	.txt	A TSV (tab-separated values) file with column definitions as: 1. Zero-based index of the parcellation label (0 defines the background) 2. Red color component(int, max 255) of the RGB color representing the label 3. The green color component 4. The blue color component 5. Label transparency (float 0.00 .. 1.00) 6. Label visibility (0 or 1) 7. Label mesh visibility (0 or 1) 8. Label description

ACKNOWLEDGEMENTS
The authors would like to acknowledge the efforts of many other individuals who spent many hours examining the experimental materials under the microscope, conducting digital image analysis, or developing software components, which made this resource possible, including Dr. Nafiseh Atapour, Dr. Kathleen Burman, Lorenzo Canalini, Dr. Michela Gamberini, Dr. Daniele Impieri, Daria Malamanova, Susan Palmer, Jan Mąka, Karyn Richardson, Maciej Śmigielski, Gabriela Saworska, Ianina H. Wolkowicz, Amanda Worthy and Sherry Zhao, Technical support from Kirsty Watkins in conducting histological processing, and the contributions of Dr. Tristan Chaplin and Dr. Hsin-Hao Yu to early phases of the development of the methods are also gratefully acknowledged. This work utilized the Multi-modal Australian ScienceS Imaging and Visualization Environment (MASSIVE) high-performance computing infrastructure (https://www.massive.org.au/), and scanning of histological slides was performed by the Monash Histology Platform (https://platforms.monash.edu/histology/). Funding for the experiments and the development of the resource was provided by the Australian Research Council (DP110101200, DP140101968, CE140100007), the National Health and Medical Research Council (APP1003906, APP1020839, APP1083152, APP1082144, APP1122220), the International Neuroinformatics Coordinating Facility (INCF Seed Funding grant scheme), the European Regional Development Fund under the Operational Program Innovative Economy (POIG.02.03.00-00-003/09), the National Centre for Research and Development (ERA-NET-NEURON/17/2017), the National Institute of Mental Health (R01MH062349), the Office of Naval Research (N00014-17-1-2041), Cold Spring Harbor Laboratory (Crick-Clay Professorship), the Mathers Foundation, and the Indian Institute of Technology Madras H. N. Mahabala Chair).

AUTHOR CONTRIBUTIONS
Piotr Majka: methodology, software, validation, data analysis, data curation, visualization, supervision, project administration
Shi Bai: software, visualization
Sophia Bakola: data analysis
Sylwia Bednarek: software
Jonathan M. Chan: data analysis
Natalia Jermakow: data analysis
Lauretta Passarelli: data analysis
David H. Reser: data analysis
Panagiota Theodoni: data analysis
Katrina H. Worthy: data analysis
Xiao-Jing Wang: supervision, project administration
Daniel K. Wójcik: methodology, supervision, project administration, funding acquisition
Partha P. Mitra: conceptualization, methodology, supervision, project administration
Marcello G. P. Rosa: conceptualization, methodology, validation, data analysis, data curation, supervision

REFERENCES

Avants, B. B. *et al.* A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage* **54**, 2033–2044 (2010), doi: 10.1016/j.neuroimage.2010.09.025.

Majka, P. *et al.* Towards a comprehensive atlas of cortical connections in a primate brain: Mapping tracer injection studies of the common marmoset into a reference digital template. *The Journal of Comparative Neurology* **524**, 2161–2181 (2016), doi: 10.1002/cne.24023.

Majka, P., Wójcik, D.K. Possum – A Framework for Three-Dimensional Reconstruction of Brain Images from Serial Sections. *Neuroinformatics* **14**, 265–278 (2016), doi: 10.1007/s12021-015-9286-1.

Majka, P. *et al.* Open access resource for cellular-resolution analyses of corticocortical connectivity in the marmoset monkey. *Nature Communications* **11**, (2020), doi: 10.1038/s41467-020-14858-0.

Paxinos, G., Watson, C., Petrides, M., Rosa, M. & Tokuno, H. The Marmoset Brain in Stereotaxic Coordinates. Elsevier Academic Press (2012), ISBN: 978-0-12-415818-4.