



OPEN

DATA DESCRIPTOR

A three-dimensional thalamocortical dataset for characterizing brain heterogeneity

Judy A. Prasad¹, Aishwarya H. Balwani², Erik C. Johnson³, Joseph D. Miano⁴, Vandana Sampathkumar¹, Vincent De Andrade⁵, Kamel Fezzaa⁵, Ming Du⁵, Rafael Vescovi⁵, Chris Jacobsen^{5,6}, Konrad P. Kording⁷, Doga Gürsoy⁵, William Gray Roncal³, Narayanan Kasthuri¹ & Eva L. Dyer^{2,8} ✉

Neural microarchitecture is heterogeneous, varying both across and within brain regions. The consistent identification of regions of interest is one of the most critical aspects in examining neurocircuitry, as these structures serve as the vital landmarks with which to map brain pathways. Access to continuous, three-dimensional volumes that span multiple brain areas not only provides richer context for identifying such landmarks, but also enables a deeper probing of the microstructures within. Here, we describe a three-dimensional X-ray microtomography imaging dataset of a well-known and validated thalamocortical sample, encompassing a range of cortical and subcortical structures from the mouse brain. In doing so, we provide the field with access to a micron-scale anatomical imaging dataset ideal for studying heterogeneity of neural structure.

Background & Summary

Whether focusing on a large swath of cortex or a single subcortical nucleus, consistent and reliable visualization of brain microarchitecture is critical for the creation of reference points which demarcate the brain's landscape¹. This is true not only for the identification of landmarks (or regions of interest), but also the study of local circuits therein. Thus, detailed views into the brain's microarchitecture can be used to study disease, experimentally target circuits, and to advance the field's understanding and integration of each of these overarching neural systems.

With advances in the reconstruction and analysis of significantly larger brain volumes, neuroscientists are now able to visualize patterns of microarchitecture that arise at a scale previously inaccessible using traditional methods^{2–4}. Examples such as CLARITY⁵, expansion microscopy⁶, serial two photon tomography^{7–9}, multi-beam scanning electron microscopy¹⁰, and X-ray microtomography^{11–15}, now provide access to several regions of interest within a volume of tissue simultaneously, providing rich context to study both local circuitry and long-range projections. With many of these new techniques, it is possible to image and analyze large intact anatomical samples that preserve the connectivity between multiple regions of interest^{16,17}, thus providing a lens into the heterogeneity of neural structure within and across different brain areas.

Here, we introduce a three-dimensional neuroanatomical dataset extracted from a validated, in-vitro mouse thalamocortical sample spanning six anatomically distinct regions of interest (somatosensory cortex, two thalamic nuclei, zona incerta, striatum and hypothalamus)¹⁶. This dataset was reconstructed using X-ray microtomography to reveal a diverse composition of microstructures (e.g., myelinated axons, cell bodies, and vasculature) within each region at isotropic, micron-scale resolution. For a selected number of regions, we also provide validated annotations performed at the area level (which identifies regions of interest several hundreds of microns in size) and at the pixel level (which identifies microstructures at a micron scale). To technically validate the dataset, human annotators assessed two series of extracted images. We found that both annotators were able to classify

¹Department of Neurobiology, University of Chicago, Chicago, IL, USA. ²School of Electrical & Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ³Applied Physics Laboratory, Johns Hopkins University, Laurel, MD, USA. ⁴College of Computing, Georgia Institute of Technology, Atlanta, GA, USA. ⁵Advanced Photon Source, Argonne National Laboratory, Argonne, IL, USA. ⁶Department of Physics & Astronomy, Northwestern University, Evanston, IL, USA. ⁷Department of Biomedical Engineering, University of Pennsylvania, Philadelphia, PA, USA. ⁸Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ✉e-mail: evadye@gatech.edu

images from these datasets reliably and accurately, likely in part due to the heterogeneity of the microstructures throughout each regions of interest. This dataset and multi-scale annotations are available for visualization and download from an online, interactive atlas (<http://bossdb.org/project/prasad2020>).

Ultimately we envision this heterogeneous, 3D brain volume dataset as a resource not only for neuroscientists interested in exploring structures within this thalamocortical pathway, but also machine learning scientists seeking data diverse enough to test computer vision methods for brain area prediction and segmentation. We further believe the provision of this dataset will prompt collaborative opportunities for both experimentalists and theorists interested in exploring neural circuitry at the micron-level and beyond.

Methods

Sample preparation. All animal experiments were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Chicago. The thalamocortical sample for this dataset was obtained from an 8 week old, C57BL/6J female mouse. The animal was deeply anesthetized using Euthasol (60mg/kg), then transcardially perfused. Vasculature was first flushed with 0.1M cacodylate buffer, followed by primary fixatives paraformaldehyde (2%) and glutaraldehyde (2.5%) in 0.1M cacodylate buffer. The brain was dissected from the skull and then post-fixed for 48 hours. Following multiple rinses in 0.1M cacodylate buffer, the brain was sliced on a vibratome at a thickness of 450 μm until the thalamocortical slice was obtained¹⁶. At this point, a 1.7 by 6.5 mm strip of tissue which preserves the pathway from somatosensory cortex to the ventral posterior thalamic nucleus was dissected (see Fig. 1a–c)¹⁶. Prior to imaging, the total estimated volume of this sample was 5 mm³. The sample was further post-fixed in paraformaldehyde (2%) and glutaraldehyde (2.5%) for 2 hours at room temperature, rinsed three times with 0.1M cacodylate buffer, and stored overnight in 0.1M cacodylate buffer at 4° Celsius. The tissue was then embedded with heavy metals as described by¹⁸. The sample was initially stained with (2%) buffered osmium tetroxide for 1.5 hours at room temperature followed by (2.5%) potassium ferrocyanide for 1.5 hours at room temperature. After rinsing with water, tissue was incubated in (1%) filtered thiocarbohydrazide at 40 degrees Celsius for 45 minutes. Following rinsing with water, the tissue was stained with another round of (2%) unbuffered osmium tetroxide for 1.5 hours at room temperature. After another thorough rinse with water, the sample was stained with (1%) aqueous uranyl acetate overnight at 4 degrees Celsius and at 50 degrees Celsius for 2 hours. Following the final water rinse, tissue was stained with lead aspartate for 2 hours at 50 degrees Celsius. This was followed by dehydration through a series of graded ethanols and propylene oxide, and a gradual infiltration of the tissue with epon resin. The infiltrated sample was incubated in 100% epon resin overnight, before being placed in 1.5 mm cylindrical tubing with fresh 100% resin. The preparation was then cured in an oven at 60° Celsius for 48–72 hours.

Imaging and reconstruction. Synchrotron X-ray tomography was performed on the embedded sample on the 32-ID beamline at the Advanced Photon Source in Argonne National Laboratory as described by Vescovi and colleagues¹⁹. X-ray radiographs were recorded with a detection system consisting of a LuAG:Ce scintillator converting X-rays into visible light that were magnified with a 5X objective lens onto a CCD detector with 1920 x 1200 pixels (5.86 μm ; Flir Grasshopper 3, Model #GS3-U3-23S6M-C). To improve the spatial resolution, the detector was built with a large NA (0.21) long working distance Mitutoyo 5X objective lens with a resolving power of 1.3 μm and 14 μm depth of focus. To maintain its resolving power, the lens is coupled with a 13 μm thick, thin film scintillator matching its depth of focus²⁰. With a 5X magnification, the pixel size of each projection image was 1.17 μm . Exposure for a single projection image took approximately 30ms, thus the total imaging acquisition time for the collection of 3000 projections was approximately 1.5 min in total (see Figure 1d for an illustration of the imaging process).

Each single reconstructed dataset corresponded to a region of 1920 × 1920 × 1200 voxels. These volumes were then stitched together using Tomosaic software¹⁹. The entire volume was trimmed down to 720 × 1420 × 5805 voxels³ which corresponded to 0.842 × 1.661 × 6.792 mm³. Acquired data was stored in HDF5 files with Dexchange format (see²¹). To make subsequent analysis and labeling more aligned with an anatomical frame of reference, we virtually resliced the data to produce 720 images (each 1420 × 5805 pixels), where each image spans all cortical layers and traversed all sub-cortical regions of interest as well. This resulted in an image volume of 5.9 Gigavoxels total, including pixels outside of the sample. These images were chunked into 14 independent stacks comprised of 50 .tiff files, spanning the entire length of the sample. These datasets were converted to 8 bits from 32 bit precision after histogram normalization.

Ground truthing and annotation for brain areas (regions of interest). 6 regions of interest were annotated by an experienced neuroanatomist using a macroscale view of the entire thalamocortical sample (from somatosensory cortex to hypothalamus). 9 of the above 14 stacks were used for annotations, performed using the 3D segmentation software ITK-Snap²². Nine images (slices in Z) distributed uniformly throughout the volume ($z = z_1, z_2, \dots$) were annotated at the pixel level for each region of interest, approximately 50 images apart. The regions annotated included: somatosensory cortex (CTX), striatum (STR), the thalamic reticular nucleus (TRN), the ventral posterior (VP) nucleus of thalamus, zona incerta (ZI), the hypothalamus (HYP), and white matter (WM), which includes the internal capsule and corpus callosum. Each region's annotation was therefore spaced 50 slices (58.5 μm) apart throughout most of the volume ($z = 109, 159, 209, 259, 309, 359, 409, 459$), with the exception of the HYP which was only present for the first four stacks within the dataset. The span of the volume considered was bounded by slices 109 to 459, as regions beyond these slices were insufficiently visualized to provide complete annotations. It took approximately 2 hours to complete annotations for all regions of interest across the 14 stacks.

Pixel-level annotations of selected regions of interest. We manually created ground truth pixel-level annotations of 4 major brain areas within the dataset: CTX, ZI, STR, and TRN. These regions were selected as they

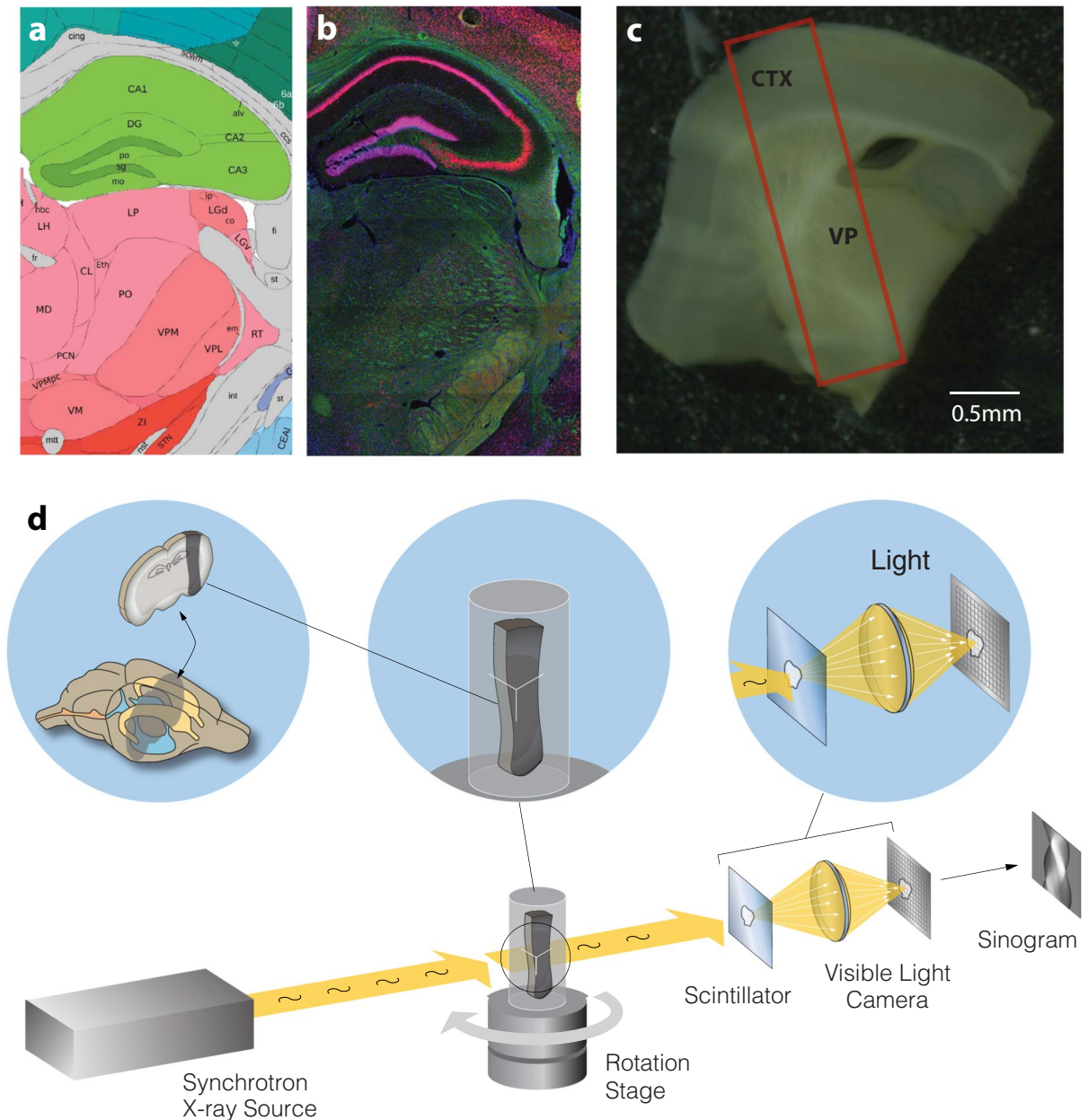


Fig. 1 Overview of the thalamocortical sample microarchitecture and 3D reconstruction. Data from the Allen Reference Atlas (ARA) (<http://mouse.brain-map.org/static/atlas>) provides a schematic overview of the dataset regions of interest (a), along with cytoarchitectural differences as identified by NeuN staining (b)⁸. (c) The photomicrograph to the right shows the thalamocortical slice prior to dissection, with the final sample volume outlined in red. CTX = somatosensory cortex; VP = ventral posterior nucleus (d) Visualization of the synchrotron X-ray microtomographic data acquisition process. X-ray projections were acquired and reconstructed into a 3D image volume with micron-scale isotropic resolution (1.17 μm pixel size).

spanned the full extent of the sample, beyond simply cortical and thalamic regions of interest (i.e., through the inclusion of STR and ZI), making them well suited to studying architectural uniqueness within and across brain regions. To standardize the ground truth segmentations, we extracted 4 volumes of size (x:y:z) = (257:257:361), 1 for each of the 4 brain areas. The coordinates of each of these volumes are as follows, formatted as (xstart_xend_ystart_yend_zstart_zend): CTX (4600_4857__900_1157__110_471), ZI (1543_1800__650_907__110_471), STR (3700_3957__500_757__110_471), TRN (3063_3320__850_1107__110_471). Within each of these (257:257:361) brain area volumes, we densely annotated (starting at index z = 0) slice z = 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, 330. This results in 11 densely annotated images per region and 44 images in total across the four areas, which took a total of 44 hours to complete .

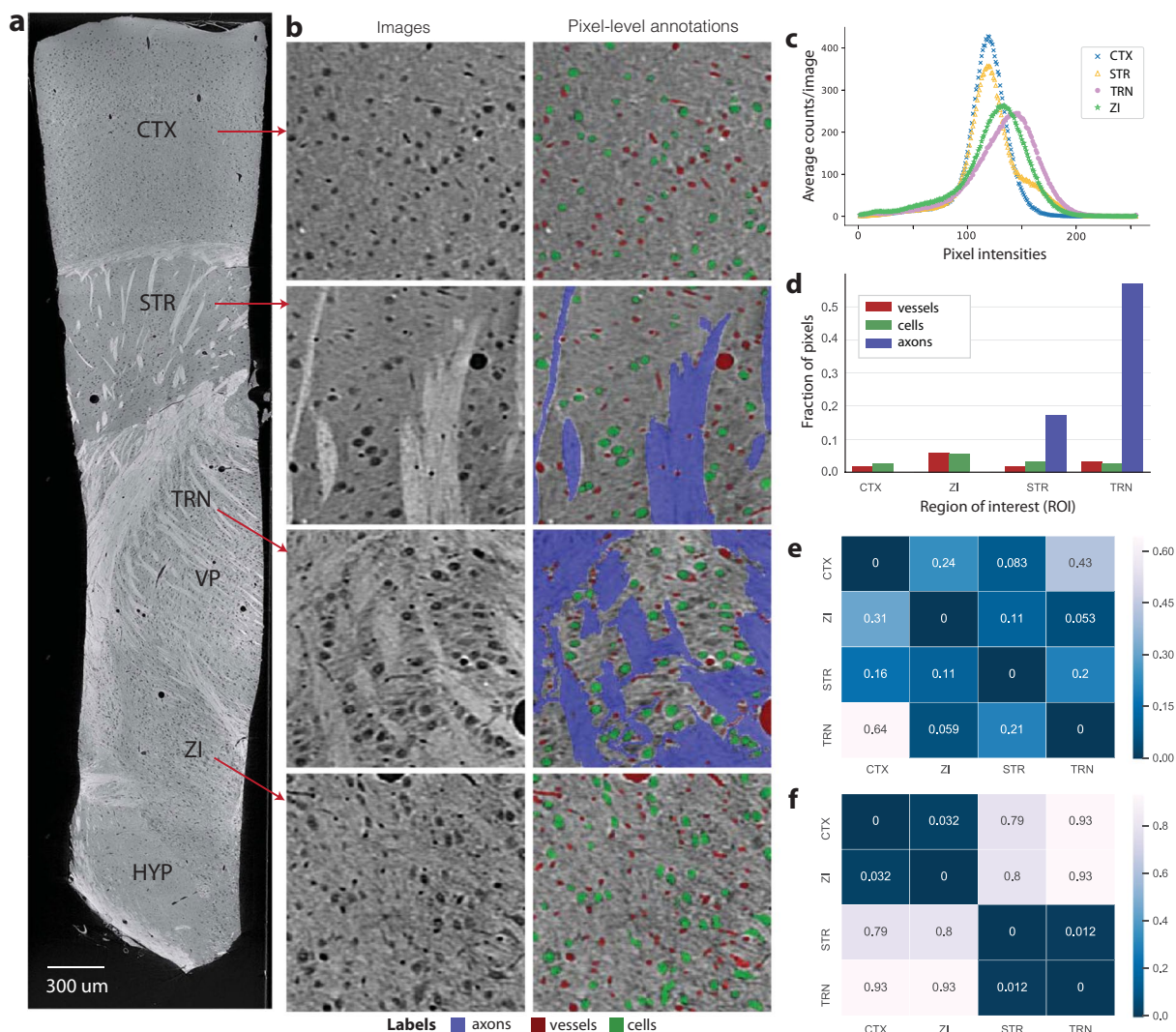


Fig. 2 Validation of neuroanatomical heterogeneity within the dataset. In (a), an example of a reconstructed image from the dataset following X-ray acquisition, highlighting the regions of interest in the sample. From top to bottom: somatosensory cortex (CTX); striatum (STR); the thalamic reticular nucleus (TRN) and the ventral posterior nucleus (VP) of thalamus; zona incerta (ZI); and hypothalamus (HYP). (b) Examples of the microstructures identified manually within the different ROIs, including cells, axons and blood vessels. These examples each span a roughly 300×300 micron field-of-view and highlight the architectural diversity within and across regions of interest. (c) The distribution of pixel intensities across four selected regions of interest within the dataset (CTX, STR, TRN, ZI). (d) The distribution of pixels divided by underlying microstructure class (cell, blood vessel, axon) within each region of interest. In (e,f), we show the KL-divergence between: the pixel intensity distributions across the selected regions (e), and the microstructural composition of selected regions as measured with dense manual annotations (f).

Data Records

The X-ray imaged dataset and associated annotations were deposited in figshare^{23–25} and in bossDB (<http://bossdb.org/project/prasad2020>). bossDB is a spatial database that is optimized for access and visualization of three-dimensional neuroimaging data, allowing users to efficiently and dynamically access different sub-volumes of the data²⁶. This database connects seamlessly to Neuroglancer²⁷ which enables interactive visualization of large-scale, volumetric images, annotations, and analytics results from a web browser. Within this web portal, users can view and navigate the data in three dimensions using public access credentials (no account creation required). In addition to these web-based tools, we provide example Jupyter notebooks to demonstrate approaches for downloading raw data and annotations, and applying simple analysis algorithms to subvolumes of data (<http://nerdslab.github.io/xray-thc>). These tools directly address the needs of both novice and expert users and are easily adapted to additional use-cases.

The stored data and annotations consist of the following three sets of images and annotations:

- *Original Image Data*²³: The raw images are rescaled from 32-bits and stored in 8-bit format. We computed the average number of bits of information in each pixel of the original image, confirming that an 8-bit depth was

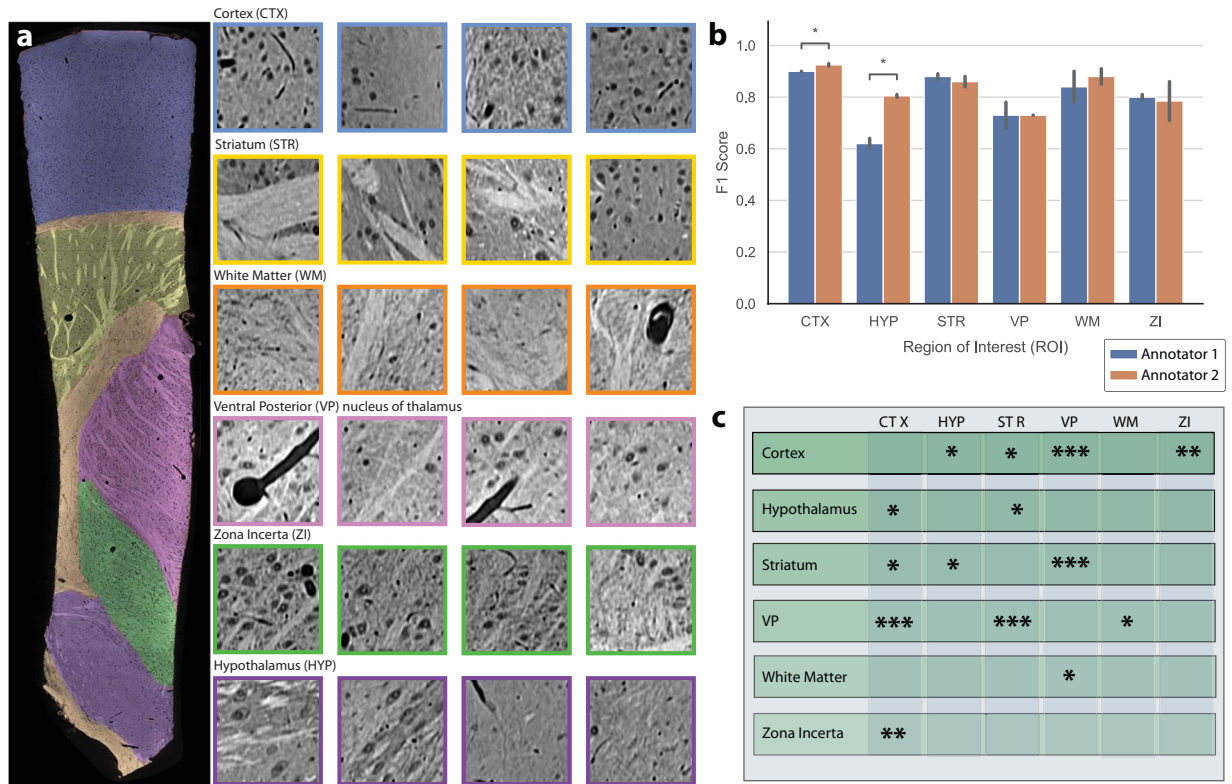


Fig. 3 Brain area prediction performance. **(a)** An annotated image from the dataset with each region of interest overlaid as a distinct color (left) and 150 × 150 micron snapshots from within (right), highlighting microstructural heterogeneity within each region. **(b)** The performance (f1-score) and inter-rater reliability of two annotators classifying image patches similar to those visualized in **(a)**. Both annotators classified images into one of six different brain areas; each test set consisted of 180 images (30/class) for a total of 360 images classified. **(c)** Summary of significance in annotators' ability to accurately predict a region of interest relative to others. Asterisks denote prediction measures that are significantly different, where * is used to denote $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$.

sufficient to capture the information in the CT stack. These images were then stacked into an image volume of size $5805 \times 1420 \times 720$ (x,y,z). The pixel size is $1.17 \mu\text{m}$ isotropic. Valid labels span the entire X, Y, and Z axis of the volume. Slices spanning from $z = 15$ to $z = 670$ contain brain specimen, and outside of this range, the epon block the sample is embedded in is visible. In bossDB, these annotations are stored in a channel called images.

- **Region-of-interest (ROI) Annotations²⁴**: Manually labeled pixel-level annotation of brain regions of interest. Valid labels span the entire X and Y axis of the volume. Valid z slices are 109, 159, 209, 259, 309, 359, 409, 459. The labels are 0-> no label; 1-> cortex; 2-> striatum; 3-> trn; 4-> vp; 5-> zona incerta; 6-> internal capsule; 7-> hypothalamus; 8-> corpus callosum. In bossDB, these annotations are stored in a channel called region_of_interest.
- **Pixel-level Microstructure Annotations²⁵**: Manually labeled pixel-level annotation of neural microstructure. Four volumes of size $361 \times 257 \times 257$ are annotated. The first volume, from cortex, spans z from 110 to 471, y from 900 to 1157, and x from 4600 to 4857. The second volume, from striatum, spans z from 110 to 471, y from 500 to 757, and x from 3700 to 3957. The third volume, from vp, spans z from 110 to 471, y from 850 to 1107, and x from 3063 to 3320. The fourth volume, from zona incerta, spans z from 110 to 471, y from 650 to 907, and x from 1543 to 1800. The labels are 0-> no label (background); 1-> vasculature; 2-> cell body; 3-> myelinated axon. In bossDB, these annotations are stored in a channel called pixel_annotation.

Technical Validation

Examination of pixel-level features across regions of interest. As one of the defining features of this thalamocortical dataset is its diverse collection of brain regions, we sought to assess the extent of tissue heterogeneity by examining four major brain areas at a microstructural level (CTX, ZI, STR, TRN). Specifically, the intensities of pixels within each region and the anatomical composition (cells, blood vessels, and axons) were measured (see Fig. 2c,d). We determined the composition of features for each region (assessing the percentage of blood vessels, cells and myelinated axons within each), based on manual annotations over 128 images (32/class, $150 \mu\text{m}$ by $150 \mu\text{m}$ in size). While relatively similar numbers of cells and vasculature were identified within each region of interest, the fraction of axons annotated within each region varied dramatically, with far fewer axons identified in

the CTX and ZI relative to a high proportion of axons in the STR and TRN (Fig. 2d). The KL divergence between pixel intensity distributions (Fig. 2e) revealed that CTX is most dissimilar from TRN, which is evident given the vast differences in microstructure between these two regions of interest. Manual annotations further confirmed TRN and CTX were highly dissimilar in their microstructural composition, whereas the TRN and STR were highly similar as measured by the KullbackLeibler (KL) divergence (Fig. 2f). One unexpected finding was, in spite of the difference in their microstructural composition, STR and CTX were highly similar in their pixel intensity distributions. The 3D reconstruction of our thalamocortical dataset with microCT thus provided richer details that are unobservable upon examining stained, thinly sliced tissue using traditional light microscopy methods.

Region of interest prediction from local views of the microarchitecture. This dataset is comprised of a range of brain regions which can be visualized and annotated by any user interested in exploring macro-level or local characteristics. To validate the use of this dataset with annotators, we assessed whether humans can accurately predict regions of interest within the sample using only a small field-of-view (150x150 microns); see Fig. 3a. For simplicity, TRN and VP were combined into a single region of interest (VP). The internal capsule and corpus callosum fiber tracts were included in this study, and categorized as WM. We then provided a training set of 48 images (8 per region of interest) for 2 annotators to examine, which took approximately less than 10 minutes for the annotators to get acquainted. After studying these examples, each annotator was provided with a test set of 180 novel images (30 per region of interest) to sort into one of six region of interest categories. This sorting procedure took both annotators less than 2 hours to complete. These annotators had previously studied example images from each region of interest prior to classifying these images, without extensive training. Interestingly, both annotators classified images within the CTX, STR and WM to a high degree of accuracy (>80%), whereas images from HYP, VP and ZI proved more challenging to classify (see Fig. 3b). Generally, CTX images were classified to the same degree of accuracy as WM images relative to other regions of interest (see Fig. 3c). We also noted that the annotators themselves varied in their classification performance; while Annotator 2 was slightly more accurate at classifying images from CTX relative to Annotator 1, they both had significantly more difficulty with identifying images from the HYP (* $p < 0.05$). It was also evident that the images from HYP were most challenging to classify, regardless of annotator and relative to those from CTX and STR (which had highly distinguishing microstructures, and therefore more likely to be correctly classified). Collectively, these findings support this heterogeneous 3D imaging dataset serving as a generalizable and useful resource for the field, given that human annotators can use it to a high degree of success for image classification.

Code availability

Code for downloading the data and annotations in bossDB can be found in the 'data_access_notebooks' folder here: <https://github.com/nerdslab/xray-thc>. A Jupyter notebook for generating the results in Figs. 2, 3 can be found in the 'analysis_notebooks' folder in the same repo. Annotations, images, and analysis notebooks used for the inter-rater reliability study, are also provided through figshare to facilitate reproducibility²⁸. All of these examples are written in Python 3 and executed using Jupyter notebooks, a cross platform Python solution.

Received: 20 April 2020; Accepted: 17 September 2020;

Published online: 20 October 2020

References

- Schmitz, C. & Hof, P. Design-based stereology in neuroscience. *Neuroscience* **130**, 813–831 (2005).
- Lein, E. S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
- Tomer, R., Ye, L., Hsueh, B. & Deisseroth, K. Advanced clarity for rapid and high-resolution imaging of intact tissues. *Nature protocols* **9**, 1682 (2014).
- Rolnick, D. & Dyer, E. L. Generative models and abstractions for largescale neuroanatomy datasets. *Current opinion in neurobiology* **55**, 112–120 (2019).
- Ueda, H. R. *et al.* Tissue clearing and its applications in neuroscience. *Nature Reviews Neuroscience* **21**, 61–79 (2020).
- Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).
- Ragan, T. *et al.* Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nature methods* **9**, 255–258 (2012).
- Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
- Harris, J. A. *et al.* Hierarchical organization of cortical and thalamic connectivity. *Nature* **575**, 195–202 (2019).
- Xu, C. S. *et al.* A connectome of the adult drosophila central brain. Preprint at <https://doi.org/10.1101/2020.01.21.911859v1> (2020).
- Mizutani, R. *et al.* Three-dimensional x-ray visualization of axonal tracts in mouse brain hemisphere. *Scientific reports* **6**, 35061 (2016).
- Dyer, E. L. *et al.* Quantifying mesoscale neuroanatomy using x-ray microtomography. *eNeuro* **4**, 1–18 (2017).
- Töpperwien, M., van der Meer, F., Stadelmann, C. & Salditt, T. Threedimensional virtual histology of human cerebellum by x-ray phase-contrast tomography. *Proceedings of the National Academy of Sciences* **115**, 6940–6945 (2018).
- Busse, M. *et al.* Three-dimensional virtual histology enabled through cytoplasm-specific x-ray stain for microscopic and nanoscopic computed tomography. *Proceedings of the National Academy of Sciences* **115**, 2293–2298 (2018).
- Itokawa, M. *et al.* Cutting-edge morphological studies of post-mortem brains of patients with schizophrenia and potential applications of x-ray nanotomography (nano-ct). *Psychiatry and clinical neurosciences* **74**, 176–182 (2020).
- Agmon, A. & Connors, B. Thalamocortical responses of mouse somatosensory (barrel) cortex in vitro. *Neuroscience* **41**, 365–379 (1991).
- Llano, D. A., Slater, B. J., Lesicko, A. M. & Stebbings, K. A. An auditory colliculothalamocortical brain slice preparation in mouse. *Journal of neurophysiology* **111**, 197–207 (2014).
- Hua, Y., Laserstein, P. & Helmstaedter, M. Large-volume en-bloc staining for electron microscopy-based connectomics. *Nature communications* **6**, 1–7 (2015).
- Vescovi, R. *et al.* Tomosaic: efficient acquisition and reconstruction of teravoxel tomography data using limited-size synchrotron x-ray beams. *Journal of synchrotron radiation* **25**, 1478–1489 (2018).
- Douissard, P.-A. *et al.* Epitaxial growth of luag: Ce and luag: Ce, pr films and their scintillation properties. *IEEE Transactions on Nuclear Science* **63**, 1726–1732 (2016).

21. De Carlo, F. *et al.* Scientific data exchange: a schema for hdf5-based storage of raw and analyzed data. *Journal of synchrotron radiation* **21**, 1224–1230 (2014).
22. Yushkevich, P. A. *et al.* User-guided 3d active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *Neuroimage* **31**, 1116–1128 (2006).
23. Prasad, J. A. *et al.* A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: X-ray microCT images (Tiff). *figshare* <https://doi.org/10.6084/m9.figshare.12153573.v1> (2020).
24. Prasad, J. A. *et al.* A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: Region of Interest Annotations (Nrrd). *figshare* <https://doi.org/10.6084/m9.figshare.12153549.v1> (2020).
25. Prasad, J. A. *et al.* A three-dimensional thalamocortical dataset for characterizing brain heterogeneity: Microstructure Annotations (NumPy). *figshare* <https://doi.org/10.6084/m9.figshare.12153516.v1> (2020).
26. JHU/APL. bossDB-Intern. <https://github.com/jhuapl-boss/intern> (2020).
27. Google. Neuroglancer. <https://github.com/google/neuroglancer> (2020).
28. Balwani, A., Miano, J., Prasad, J. & Dyer, E. Multi-scale annotations of neural structure in a thalamocortical sample spanning many brain regions. *figshare* <https://doi.org/10.6084/m9.figshare.12153498.v2> (2020).

Acknowledgements

E.L.D., A.H.B., and J.D.M. were supported by award NSF IIS-1755871. E.L.D., E.J., and W.G.R. were supported by NIMH grant R24MH114799. Support for bossDB comes from NIMH grant R24MH114785. M.D. and C.J. were supported by NIMH awards U01 MH109100 and R01 MH115265. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. We thank Michelle Alba for her assistance with the brain area prediction experiments described above.

Author contributions

J.A.P., E.L.D., K.P.K. and N.K. conceived of the dataset. J.A.P. and E.L.D. designed the research and wrote the paper in collaboration with all co-authors. V.S. perfused the animal and processed the tissue for imaging; J.A.P. dissected the thalamocortical sample. V.D.A., K.F., M.D., R.V., C.J., D.G. collected and/or generated the X-ray imaging data. V.D.A. and K.F. created the custom imaging setup for microCT used to collect this data at beamline 32-ID at the Advanced Photon Source. J.A.P. performed annotation of all area-level ROIs. J.D.M. provided all pixel-level annotations, curated the ground truth datasets and image volumes, and helped develop code to access subvolumes and annotations using the Bossdb API. A.H.B. conducted all aspects of the human annotation experiment (including the extraction of training datasets and analysis of annotator performance), analyzed and validated the data. E.C.J. and W.G.R. ingested data into bossDB and designed code to pull down and interact with data in bossDB and Neuroglancer.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.L.D.

Reprints and permissions information is available at www.nature.com/reprints.

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2020