The Open University of Israel
Department of Mathematics and Computer Science

Automated, staining free cell counting in 537 murine brains discovers sex- and strain-dependent neuroanatomical features

By
David Elkind

Thesis submitted as fulfillment of the requirements toward an M.Sc. degree in Computer Science.
The Open university of Israel
Computer Science Division

Prepared under the supervision of
Prof. Noam Shental, Open University
Prof. Amit Zeisel, Technion
March, 2022

Table of Contents

List of Figures 3

List of Tables 3

Abstract 4

Background 5

Overview 5

Machine Learning 5

Deep Neural Networks 5

Feed forward networks 6

Fully connected networks 6

Convolutional neural networks 6

Image classification by a DNN 6

ResNet 7

Segmenting cells in mouse brain section images - instance segmentation 8

Mask R-CNN 8

First Stage 8

Second Stage 8

Introduction 10

Methods 11

Data 11

Training a deep neural network for cell segmentation 11

Model training and validation 11

Brain-wide automatic segmentation 12

Extracting cell information per section 12

Assigning cells to regions 13

Estimating volumes, 3D densities, and cell counts 13

Discarding whole brains or particular regions of lower technical quality 14

Results 16

Autofluorescence of STPT images display cell nuclei 16

Overview of brain-wide, regionally resolved quantification of cell density, volume and count 16

A resource for exploring neuroanatomical features across regions and populations 19

Strain differences in volume and density 22

Region-wise correlations between volume and density across brains 23

Inter-brain similarity between regions based on volume and density 24

Discussion 26

Bibliography 28

***Supplementary Material*.....30**

***תקציר*.....51**

List of Figures

Figure 1. The effect of applying residual connections. 7
Figure 2. Mask R-CNN architecture[7]. 9
Figure 3. Cell projection onto the section and the relevant volume. 14
Figure 4: Survey of neuroanatomic properties of the mouse brain.....18
Figure 5. Density and nucleus diameter along cortical regions. 19
Figure 6. Sexual dimorphism in C57BL/6J. 21
Figure 7. Sexual and cross-strain dimorphism in C57BL/6J (B6) and FVB.CD1 (CD1). 23
Figure 8. Correlations between volume, cell count, and density. 24
Figure 9. Visualizing the similarity between brain regions based on ABCP. 25

List of Tables

Table 1. Breakdown of the data by strain and sex.....11
Table 2. Model performance over out-of-sample tiles. 12

Abstract

The mouse brain is by far the most intensively studied among mammals, yet estimation of its cell density and variability across brains is out of reach for many regions, and for others estimates are based on extrapolations. Furthermore, the question of variation between individuals in region-specific cell density and volume is scarcely addressed in the literature. The Allen Brain Institute produces high-resolution full brain images for hundreds of brains. Although these were created for a different purpose, they can serve as a first attempt to address such questions. Here, we aim to systematically characterize cell density and volume for each anatomical unit in the mouse brain over hundreds of brains. We developed a deep neural network-based segmentation pipeline that uses the auto-fluorescence intensities of the images to segment cell nuclei even within the densest regions, such as the dentate gyrus. We applied our pipeline over 537 brains of males and females from C57BL/6J and FVB.CD1 strains to assess strain-specific and sex-specific changes. We found that increased overall brain volume does not result in uniform expansion across all regions. Moreover, region-specific density changes are often negatively correlated with the volume of the region, therefore cell count does not scale linearly with volume. Systematic brain-wide cell counting is a powerful tool for detecting variability and small differences across populations. We provide the results of this analysis as an accessible tool for the community.

Background

Overview

This thesis applies machine learning, specifically, deep neural networks to the computer vision problem of detecting cells in brain section images obtained from 537 mouse brains. The detected cells' information is subsequently used to calculate various microscopic and macroscopic properties of the processed brains and create a dataset that allows asking specific questions about properties of the mouse brain across its regions. This background shortly describes the computational tools applied as part of the thesis.

Machine Learning

Machine learning is a class of algorithms that usually perform data analysis and knowledge extraction. These algorithms are characterized by their ability to "learn", i.e., improve automatically through exposure to data samples. Machine learning can be supervised and unsupervised, where the former "learns" from labeled data, i.e., data that was assigned the desired label. In contrast, unsupervised learning "learns" from unlabeled data, usually by examining various properties of the data, such as patterns, probability distribution, etc.

The flow of a machine learning algorithm is usually centered around the "model", i.e., the entity that incorporates the parameters gathered from "learning" the data and that can be queried to produce an output for "new" data instances to which it hasn't been previously exposed. The process of learning is usually referred to as "training" whereas the process of applying the model to new data is referred to as "prediction" or "test", depending on the context.

Machine learning is applied to a wide variety of problems in a variety of disciplines. In this thesis I am applying machine learning to the computer vision problem of identifying cells in images captured using two-photon tomography.

Deep Neural Networks

Deep Neural Networks (DNN) is the most extensively used class of machine learning algorithms. Inspired by biological brains, a neural network consists of layers of artificial neurons, i.e., units that receive several inputs and produce an output which is a weighted sum of the inputs to which a mathematical function (called "activation function") is applied. The weights applied to the inputs of a neuron are the neuron's parameters, and they are calculated as a result of "training" the network by exposing it to the training data. A neural network consists of layers of neurons. Neurons from different layers can be "connected", i.e., outputs of one or more neurons from one layer serve as inputs to one or more neurons from another layer. As a minimum, a network consists of an "input" layer which represents the dimensions of the input data, and in the case of a classification task, an output layer which represents the dimensions of the label. In addition, there exist one or more internal layers referred to as "hidden layers". A deep neural network is characterized by the multiplicity of the hidden layers. There is no established convention on the number of layers that is considered "deep", however, it is common for a network to consist of tens or even hundreds of layers. The number and the size of the hidden layers, as well as all the connections between layers, are referred to as the architecture of the network.

A DNN is a function approximator, and its training process is essentially the one of numeric optimization. Correspondingly, training is implemented as a gradient descent process aimed at minimizing a loss function – the function that quantifies an error between the output predicted by the network and the ground truth. The process of gradient descent consists of repeatedly computing the loss' gradient for every neuron in the network at every training datapoint and updating its parameters so as to minimize the value of the function at the given datapoint. The process for efficient parameter updating across the network is referred to as backpropagation[1]. The loss function itself significantly impacts both the convergence of the training process and the performance of the resulting model, hence loss functions for a variety of use cases is an active research field.
The following section shorty describe the specific DNN architectures used in the thesis.

Feed forward networks

A feedforward neural network is a type of artificial neural network with no cycles formed by the connections between neurons, i.e., it is a directed acyclic graph in which the neurons are vertices, and the connections are directed edges. A feedforward neural network is the simplest type of artificial neural network. In this network, the information moves only forward, i.e., from the input nodes, through the hidden nodes to the output nodes.

Fully connected networks

A fully connected network is a feed forward network in which all neurons from a layer are connected to all neurons of the next layer. Therefore, two adjacent fully connected layers, consisting of m and n neurons respectively, form $m \times n$ connections.

Convolutional neural networks

A Convolutional Neural Network (CNN) is a feed forward network that consists of "filters" which are essentially convolution kernels that are applied to an image to form feature maps that capture the information contained in an image that is relevant to the task the networks is trained to perform[2]. One of the main characteristics of convolutional networks, that makes them especially suitable for performing computer vision tasks, is their shift invariance, i.e., the ability to recognize patterns regardless of their position in the image. Another useful trait is better computational and storage efficiency compared to the fully connected networks since the convolution filters are usually much smaller than the image itself, which leads to weight sharing between neurons in a layer.

Image classification by a DNN

Image classification is a computer vision task of classifying an image by the object that it depicts. Usually, an image contains an object of interest and possibly other objects that are not important for the classification task and are considered background. An image classification DNN usually consists of feature extraction layers and a classification head. Feature extraction layers perform dimensionality reduction by converting the image to a vector with much lesser dimensions while preserving the information essential to the specific classification task. The classification head uses

the embedding produced by the feature extraction layers to perform the actual classification. Feature extraction layers are usually implemented as a convolutional network whereas the classification head consists of a single or a small number of fully connected layers. I this thesis I am using the feature extraction layers of an image classification model as the backbone for an image segmentation task explained below.

ResNet

ResNet[3] is a CNN architecture for image classification that solves the vanishing gradient problem thereby allowing creating very deep DNNs. It has been established that the "deeper" the network (i.e., the more layers it has) the better are its prediction capabilities. However, the deeper the network the harder is training. One of the key issues complicating the training process is known as the "vanishing gradient" effect, i.e., the gradient approaches zero as it backpropagates through many layers of the network thereby making adjustment of the parameters for neurons in the layers far away from the head challenging, and hence preventing the model from converging. In fact, due to this problem, adding more layers quickly reaches the point of diminishing returns.

ResNet solves this problem by introducing "skip" connections, i.e., connections between groups of layers (which are called "residual blocks"). Hence, in addition to the regular sequential connections between adjacent layers, there are also connections that bypass certain groups of layers. Skip connections allow the gradients to "flow" freely via skip connections, thereby mitigating the vanishing gradient problem. This enables creating deeper networks while ensuring the network converges to a better accuracy during training.

As an example, Figure 1, taken from the original ResNet paper[3] depicts the training process of residual network (right) vs. a plain network (left). The latter's error does not decrease when increasing the number of layers from 18 to 34, while ResNet provides lower error using 34 layers compared to merely 18 layers.

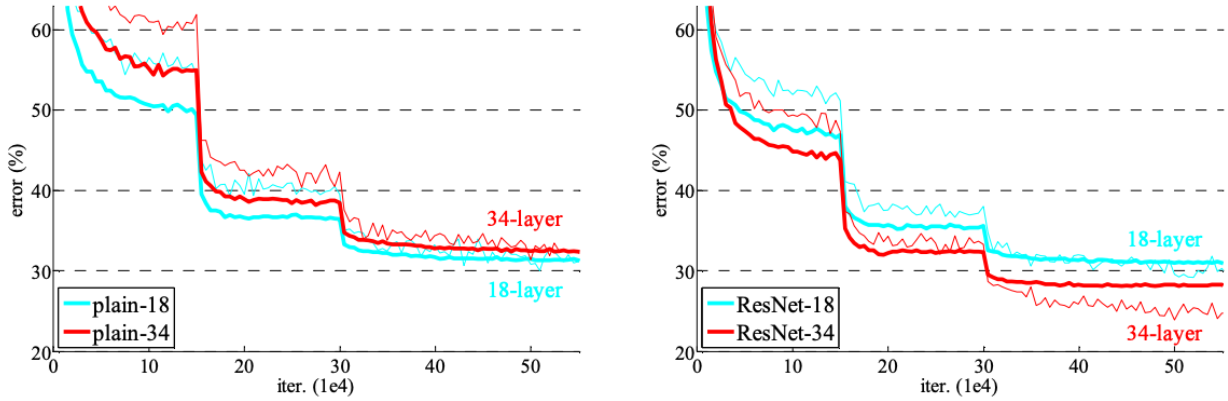

Figure 1. The effect of applying residual connections.

Performance of a regular network (left) is compared to a ResNet (right) over an image classification task of ImageNet4. Thin curves denote the training error, and bold curves denote validation error of the center crops. Left: plain networks of 18 and 34 layers. Right: ResNets of 18 and 34 layers. In this plot, the residual networks have no extra parameters compared to their plain counterparts.

Segmenting cells in mouse brain section images - instance segmentation

Instance segmentation is a computer vision task of finding instances of certain objects in an image and detecting their per-pixel segmentation mask. In my research I use instance segmentation in order to detect cells along with precise contours of their nuclei.

Mask R-CNN

Mask R-CNN[5] is the instance segmentation DNN model I am using in this research. It detects objects in the image and outputs the predicted object class with its respective confidence score and a bounding box along with the segmentation mask for each detected object. From the architecture standpoint, Mask R-CNN is a two-stage region-based convolutional network as explained below. Figure 2 depicts the Mask R-CNN architecture. In my research I used the Mask R-CNN implementation from the Detectron2[6] library. The network comprises two "stages" shortly described below.

First Stage

At the first stage, the image is being processed by (i) the feature extraction backbone. Mask R-CNN uses feature extraction layers from an image classification CNN as its backbone. Specifically, I am using ResNet-101 as the backbone. The resulting feature map is fed to the region proposal network that extracts variable size patches (regions). Each region is passed through a lightweight binary classifier which produces confidence scores to whether a region contains an object. Regions with high enough confidence scores are passed to the non-max suppressor that filters out the regions with low intersection over union for an object, so only the boxes which outline the object tightly enough are considered regions of interest passed to the predictors in the Stage 2 for further processing. Size variety of the detected regions of interest ensures that Mask R-CNN is size-independent, i.e., it is capable to detect the object regardless to their size in the image plane.

Second Stage

At the second stage, the proposed regions of interest are passed through the predictors in order to determine a match with one of the object classes the model is trained to detect, and subsequently outline the proper bounding box and segmentation mask. This is accomplished by the 3 sub-networks ("heads") as described below.

Object Class Predictor

An object class predictor is a classifier that predicts the confidence score that estimates the probability of an object belonging to each of the classes Mask R-CNN is trained to detect.

Bounding Box Predictor

Bounding Box predictor is a regression network that refines the bounding box coordinates for a region of interest.

Segmentation Mask Classifier

Segmentation mask classifier classifies pixels within the region of interest as either belonging to the object or to the background.

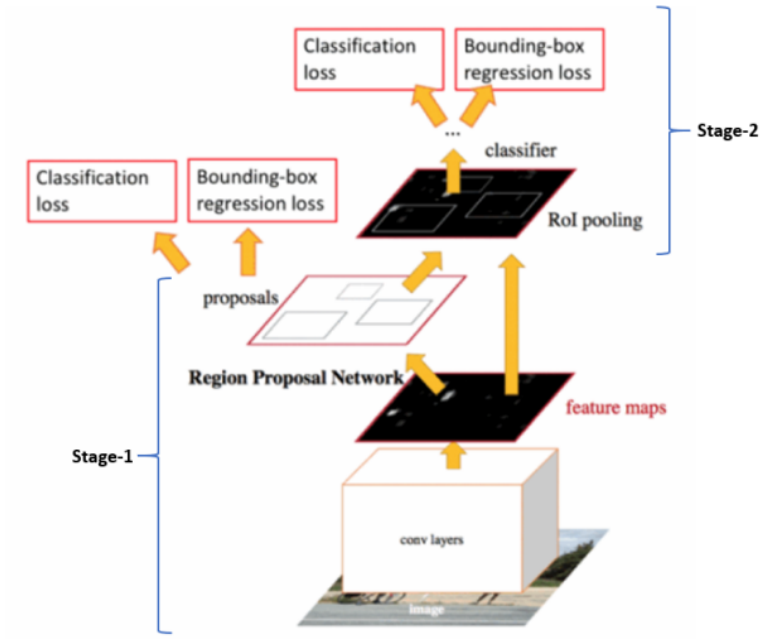

Figure 2. Mask R-CNN architecture[7].

Introduction

The mammalian brain can be divided into neuroanatomical units (e.g. brain regions) characterized by a shared function, connectivity, developmental origin, and/or cytoarchitecture (i.e. number and density of cells it contains). The mouse brain, is the most extensively studied and well characterized in terms of its regions. Cytoarchitecture is one of the most prominent features of a brain region; nevertheless, very few studies have systematically mapped cell bodies or quantified cell densities in complete mammalian brains as compared to the early, detailed cell mapping of the nematode *C. elegans* [8].

Obtaining an accurate cell count for a brain region is technically challenging. Previous estimates relied heavily on extrapolation from manual counting of 2D sections (stereology), making cell-resolved data for subcortical regions sparse [9]. Analyzing complete brains using 2D histological sections remains labor intensive because it requires sectioning, mounting and accurate alignment to a reference atlas. Furthermore, automated cell counting proved particularly difficult in dense regions such as the hippocampal formation and the cerebellum [10]. Automated block-face imaging methods solved several of these issues and drastically improved throughput [11]. For instance, serial two-photon tomography (STPT) [12] was a technological breakthrough integrating tissue sectioning with top-view light microscopy. STPT provided high-quality imaging in an optical plane below the sectioning surface and solved many problems of section distortion and atlas alignment, further easing downstream analysis. Yet, STPT typically represents a subsample of the complete volume and some interpolation is needed.
Due to their limited throughput, histological studies cannot supply the number of analyzed brains needed to uncover potential variability between individuals, experimental conditions and populations. Complementary approaches aimed at evaluating variability, e.g. Magnetic Resonance Imaging (MRI), can measure some features, such as the volume of specific brain regions, and can even track individuals along time in a noninvasive manner. Yet, MRI lacks the accuracy needed for counting cells or cell densities. Importantly, simultaneous brain-wide analysis of regional volume *and* cell counts (or density) remains difficult, especially with throughput high enough to allow comparing two experimental populations (such as two strains, or males versus females). The technical challenge to address this problem is the need for systematic measurement of all cells over hundreds of brains from multiple experimental groups.

We address this knowledge gap using the largest existing dataset of whole brain images, produced by the Allen Mouse Brain Connectivity Project. We apply a Deep Neural Network (DNN) to discern cell nuclei, using an autofluorescence channel. This enables us to perform systematic brain-wide cell density estimation over hundreds of mouse brains. Based on the alignment to the Allen Mouse Brain Atlas (AMBA), we could simultaneously measure volume *and* density for each specific brain, for each region, over a large population. We constructed a comprehensive database that aggregates these results and provide it as an accessible resource to the community. Furthermore, we discover non-trivial relationships between densities and volumes, and gain insights into strain and sex dependent characteristics across various brain regions.

Methods

Data

The Allen Mouse Brain Connectivity Project (AMBCP) dataset [13] consists of 2,992 brains, of which we processed 537 and eventually used 399 in our analysis (the strain and sex breakdown of the brains appear in Table 1). Each brain consisted of ~140 section images captured every $100\mu m$ along the anterior-posterior axis using two-photon tomography [12]. Image resolution was $0.35\mu m$ per pixel. AMBCP post-processed section images for noise removal. Rather than using the red, green, and blue channels that display brain connectivity, we used the background channel of the images, as provided by AMBCP, without additional processing, except for converting the RGB image to grayscale.

Strain	Females	Males	Total
B6.129	9	3	12
B6.129.FVB	2	-	2
B6.C3H	2	-	2
B6.FVB	3	1	4
C57BL/6J	174	195	369
FVB.CD1(ICR)	69	69	138
N/A	7	3	10
TOTAL	266	271	537

Table 1. Breakdown of the data by strain and sex

Training a deep neural network for cell segmentation

To detect cells in an image and mark their contour, we used the Detectron2 deep neural network library [6], which relies on a Mask R-CNN image segmentation model [5] with the ResNet-101 [14] as its backbone.

Model training and validation

Training the model required 3 rounds of manual annotation and training.

Initial manual annotation of the data set and model training: We annotated cell contours manually using the VGG Image Annotator software [15]. Initially, we annotated only the hippocampus, which is relatively large and easily discernible. The hippocampus contains sub-regions of different densities, which we believed would adequately represent the variety of cell densities across the mouse brain. We manually annotated tiles of 312×312 pixels ($109 \times 109\mu m$), randomly selected from the hippocampus in 5 sections of 3 brains (55 tiles in total). We provided these tiles to the network as training data, together with basic data augmentation (e.g., rotation and brightness changes) [16].

Retraining on hippocampal sections: We then applied the trained model to detect cells on a new set of 55 randomly selected hippocampus tiles. We manually corrected the results produced by the network to create a new set of ground truth annotations. Next, we retrained the model from scratch over a combined training set of 110 tiles.

Retraining on other regions: We subsequently used the trained model to detect cells on random sections of 3 selected brains. Visual inspection enabled us to select a set of 64 tiles that displayed the least accurate results and annotate them manually.

Final training: We retrained the model from scratch on the resulting training set of 174 tiles (selected from ~15 sections of ~10 brains). The total number of cells across the training set tiles was 6,247, corresponding to 0.008% of the estimated 77 million cells in the whole brain.
Technical details: We conducted the training with a batch of size 2, a learning rate of 0.00025, with decay, using the Adam optimizer [17]. Training over 174 tiles required ~395,000 iterations, and took ~36 hours using a Linux server with 160 Intel Xeon Gold 6248 2.5GHz CPUs and a Tesla V100S-PCIE-32GB GPU.
Evaluating model performance: The training process completed when the model converged. The accuracy of the model on the training data was 99.8%, with a false negative rate of 0.4%. To evaluate model performance, we manually annotated 30 additional tiles from the isocortex, medial amygdala (MEA), hypothalamus (HY), and hippocampus (HIP) of 27 brains and compared them with model prediction (Table 2). We obtained highly accurate results, comparable to the performance over the training data, for segmentation scores such as Jaccard measure [18], F1 score (harmonic mean of precision and recall), and total errors (i.e., percentage of mislabeled pixels), as well as for detection scores such as accuracy (detected cells divided by total cells) and false positive rate (false positives divided by total cells).

		Segmentation (pixelwise) scores			Detection (cellwise) scores	
Region	# cells in the test set	Jaccard Index	F1	Total errors	Accuracy	False Positive Rate
Isocortex	192	0.982	0.991	0.002	0.962	0
MEA	115	0.975	0.987	0.001	0.962	0
HY	163	0.953	0.974	0.003	0.938	0.005
HIP	566	0.986	0.992	0.001	0.979	0.009

Table 2. Model performance over out-of-sample tiles.

Brain-wide automatic segmentation

The trained DNN was applied to 537 brains, as described in detail below.

Extracting cell information per section

We divided each section into overlapping tiles sized 312×312 pixels, with an overlap of 20 pixels on each side (thus mitigating potential artifacts close to the borders of the tiles). We then applied the trained DNN to detect cells in each tile, resulting in a cell mask (i.e., a Boolean 312×312 matrix whose entries are *true* if the corresponding pixel is part of a detected cell and *false* otherwise). Next, we stitched the tiles together using a logical OR over overlapping areas, resulting in a single cell mask per section. Subsequently, we performed contour detection to obtain the coordinates of each cell in a section, and computed the morphological properties of each cell (i.e., circumference, diameter, and area). Following this analysis step, each section image was represented by a table containing the coordinates and morphological properties of its cells.

Assigning cells to regions

We used the Allen Mouse Brain Atlas (AMBA) [19] to assign the coordinates of detected cells in each section to their corresponding brain region (Table S1). But the atlas annotation was too coarse for several regions of interest, i.e., CA1, CA2, and CA3 of the hippocampus. The common denominator of these regions was the presence of a dense and a sparse region that were not separated by the atlas (e.g., the pyramidal and stratum regions of CA1, CA2, and CA3). To provide the coordinates of these sub-regions, we defined a local measure of density referred to as cell "coverage," and used it to cluster the relevant cells into a dense and a sparse region. Briefly, in a window of 64×64 pixels centered around each cell we counted the number of pixels that belong to cells, thus assigning a local "coverage" measure (the median cell area was 80 pixels, much smaller than the window around it). We then detected the sub-regions by clustering the cells according to their "coverage" values. For example, we took the "coverage" values of all CA1 cells and used K-means clustering to split them into two clusters of high and low "coverage" values. In this way, the coordinate of each cell center was assigned to either cluster. We then drew the circumference of the sub-regions by applying a standard morphological closing operation, and discarded spurious small regions.

Estimating volumes, 3D densities, and cell counts

Until this stage, the analysis provided local, i.e., microscopic properties for each detected cell, and assigned cells to a brain region. The next step was to collect cells that belong to each region and estimate their density, the volume of the region, and the total cell count. This required calculating 3D estimates based on the relevant 2D data, using the following steps:

(1) Estimating cell density per section: We used AMBA to label the area of a given region in a section. We assumed that cells belonging to a region are equi-radius spheres whose projection on the 2D section depends on the distance between their centers and the optical plane, and on the optical depth of field (Figure 3). Hence, detected cells on a 2D section s originate from a slab whose volume is $v_s = a_s \cdot (2R + d)$, where a_s is the area of a region, R is the radius of the cells in the region, and d is the optical depth of field. Cell density per section, ρ_s, is given by dividing the number of detected cells by v_s. The value of a_s is measured by pixels whose size is $0.35\mu m$, and $d = 1.5\mu m$ [12 20]. The value of R was taken as the 90th percentile of measured cell radii in a_s. The distribution of cell radii corresponds to the "projection" of the cells on the measured section, together with the optical depth of field. Downstream results of cell count and density significantly depend of the value of R, e.g., using the 50th percentile would provide larger estimated cell counts. Yet, rank order of cell counts and densities across regions is independent of the selected value of R.

(2) Calculating region volume: AMBA provides pixel-wise region annotation for each section, making it possible to calculate the area of a region per section (which is independent of cell segmentation). The 3D volume of a region is given by the sum of region volumes between adjacent sections, estimated by the average of its areas over each section. For example, if a region appears in sections 1, 2, 3, and 4, its volume is the sum of average volumes between sections 1 and 2, 2 and 3, and 3 and 4.

(3) Calculating cell counts across adjacent sections and in total: Cell counts between the adjacent sections of each region are given by the average densities in those slides multiplied by the volume

of the region between these sections. The total cell count of a region is provided by a sum across all relevant sections.

(4) Calculating cell densities per region: The overall density of each region is given by the total cell count divided by the volume of the region.

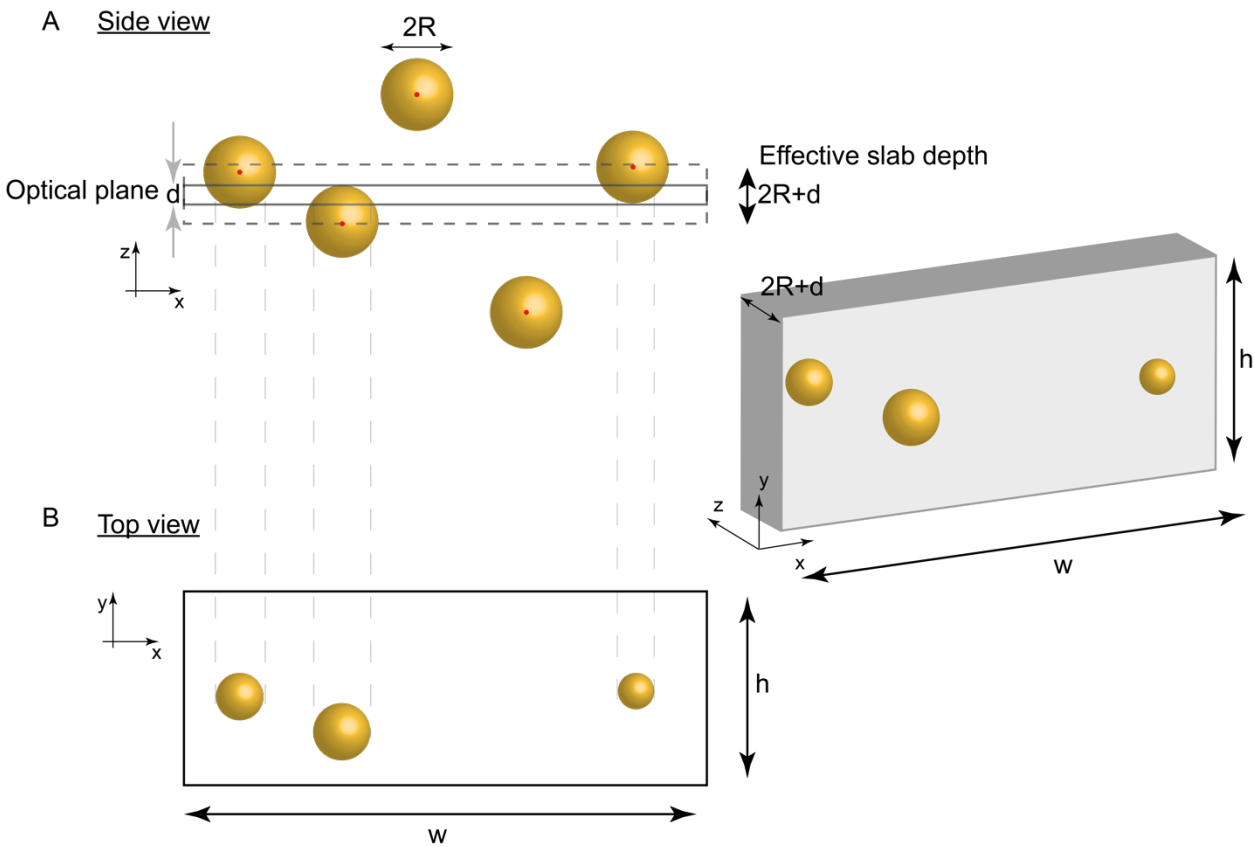

Figure 3. Cell projection onto the section and the relevant volume.

(A) Cells in 3D vs. the section plane. (B) Cell projections onto the section's plane. Cell whose center is more than $\boldsymbol{R} + \frac{\boldsymbol{d}}{\boldsymbol{2}}$ away from the plane will not be counted. (C) The resulting slab for purpose of density calculation. The slab volume is $\boldsymbol{wh(2R+d)}$ and hence the density is $\frac{\boldsymbol{3}}{\boldsymbol{wh(2R+d)}}$

Discarding whole brains or particular regions of lower technical quality

After calculating the three-dimensional counts and densities across all regions in all brains, we excluded from subsequent analysis regions and whole brains that displayed potentially flawed estimates. We applied the following criteria:

We discarded brains displaying dark images: We filtered out brains whose median brightness across the whole brain ("grey" region) was lower than 25 (on a scale between 0 and 255). In such cases, all ~140 sections of the brain were excluded from downstream analysis because DNN cell detection was either impossible or provided significantly lower estimates.

We discarded brains displaying outliers in cell count: We noticed that a common optical artifact of resolution degradation caused the DNN to falsely detect large amounts of excess cells. We marked cases in which cell count in a region was 3 standard deviations larger than the median

for the region across brains (calculated as $MAD \cdot 1.4826$, assuming normal distribution). We discarded brains that included more than three such outlier regions.

We discarded regions of small volume: We filtered out regions whose median volume across brains was smaller than $0.3mm^3$, or whose median cell count across sections was smaller than 500. We excluded such regions across all brains.

We discarded regions displaying a correlation between cell count and image brightness. We excluded regions exhibiting strong correlation (>0.25) between brightness and cell count because we assumed that in this case cell count was affected by the inability of the model to discern the cells when the brightness was too low. We discarded such regions from all brains.

We discarded regions displaying different estimates in right vs. left hemispheres: Cell count estimates in the right and left hemispheres served as a proxy for technical noise. We calculated cell counts per region using each hemisphere independently. If the difference in cell count between hemispheres for a specific brain was higher than 15.5% of the total cell count for that region, we excluded the case from downstream analysis.

Examples of excluded regions and brains appear in Figure S1. In sum, we processed 537 brains, of which 138 were fully discarded. Of 690 regions in AMBA, 369 were discarded completely. Across the remaining 399 brains and 321 regions, there were 12,016 (9%) cases in which a region was excluded.

Results

Autofluorescence of STPT images display cell nuclei

First published in 2014,[13] the Allen Mouse Brain Connectivity Project (AMBCP) project has systematically imaged 2,992 full brains using serial two-photon tomography (STPT), for the purpose of tracing neuronal projections and mapping regional (mesoscale) connectivity, using GFP-labelled viral tracers. Each brain in the dataset is covered by 130-140 (median 137) serial coronal sections, with a gap of 100μm, as reported in the AMBCP study.[21] We noticed that the red (background) channel of STPT images, taken for the purpose of atlas alignment, typically features dark, round-like objects resembling cell nuclei.

We had observed this phenomenon in our own imaging of mouse brains, but found little more than anecdotal mentions of it in the literature.[19,22,23,24] To confirm that these dark objects indeed represent cell nuclei with lower autofluorescence intensity than the surrounding lipid-rich brain tissue, we performed a standard 4% PFA perfusion-fixation followed by cryosectioning and nucleus (DAPI) counterstaining. We found the same low-autofluorescent objects, which had an overlap of nearly 100% with nuclear staining (DAPI), confirming that dark objects in STPT indeed represent cell nuclei (Suppl. Figure S2).

Overview of brain-wide, regionally resolved quantification of cell density, volume and count

To automatically collect cytoarchitecture data for each brain we trained a DNN model to detect and segment the nuclei (low-autofluorescent objects) in all brain regions, including those of the highest density, such as the dentate gyrus (DG). Due to computing constraints, we applied the model systematically to segment a subset of the AMBCP dataset comprising 537 brains (Figure 4 A-C and Methods). The model performed with an estimated 97% cell detection accuracy on a test set, with a false positive rate of <0.01 (see Methods) whenever image quality was sufficient (for exclusion criteria of whole brains or certain regions within sections, see Methods). Using detected cells in each section, we obtain a local estimate of the volumetric cell density (see Methods), that combined with the pixel-wise registration to brain regions provided by the AMBA, allow us to estimate the average cell density per region for each brain. Similarly, we evaluated the per-region volume of each brain by linear interpolation over all sections (see Methods). In sum, we simultaneously estimated the 3D cell density (D) and volume (V) of each region for each brain (see Methods). In total, we estimated per-region *D* and *V* for 532 basic regions annotated in the AMBA, which corresponds to level 6-8 of the region hierarchy.

Cell count (N) is the product $V \times D$, therefore was not considered an independent variable. The median male C57BL/6J mouse brain contained a total of 76×10^6 cells, in 367 mm^3 of grey matter, at a density of 2.05×10^5 cells/mm^3. A pie chart of the volume and cell count of the main regions (level 4 of region hierarchy) calculated across 537 brains appear in Figure 4D, and absolute cell counts for C57BL/6J male mouse representative regions are shown in Figure 4E. We quantified each level of the hierarchical tree structure of the AMBA and found good correlation (r=0.89) with a recent 3D whole-brain single-cell resolved light-sheet microscopy study[25] (Figure 4F). The diameter of detected objects (nuclei) varied between 7-9.5μm (Figure 4G left), which at a nucleus/soma volumetric ratio of 0.08[26,27] corresponds to median cell body diameters from

16.25μm in the RSPv6a, to 22μm in the ENTl3. The regional variability of cell densities was high, ranging from 1×10^5 mm^{-3} in layer 1 isocortex (e.g., MOs1) to 6×10^5 mm^{-3} in the dentate gyrus granule layer (DG-sg). We show examples of regional distributions across the full cohort of 537 brains in the inset of Figure 4G right. The large number of AMBCP brains in our analysis enabled us to compare variabilities of macroscopic properties between subsets of the cohort, e.g., to compare strains. We compared distributions of volume, cell density, and cell count at the coarsest hierarchical atlas level, i.e., across grey matter cell groups in the brains of male C57BL/6J vs. male FVB.CD1 mice (Figure 4H). Median cell density was similar for the two strains, with considerably larger variance in FVB.CD1 males. FVB.CD1, however, had 11% larger grey matter volume than C57BL/6J. Combining these two features revealed a ~10% increase in the median cell count in FVB.CD1 vs. C57BL/6J (Figure 4H right panel). These results suggest that: (*a*) there is no simple relationship between volume and density, therefore, both properties should be simultaneously measured, and (*b*) a large cohort enables detection of relatively small differences.

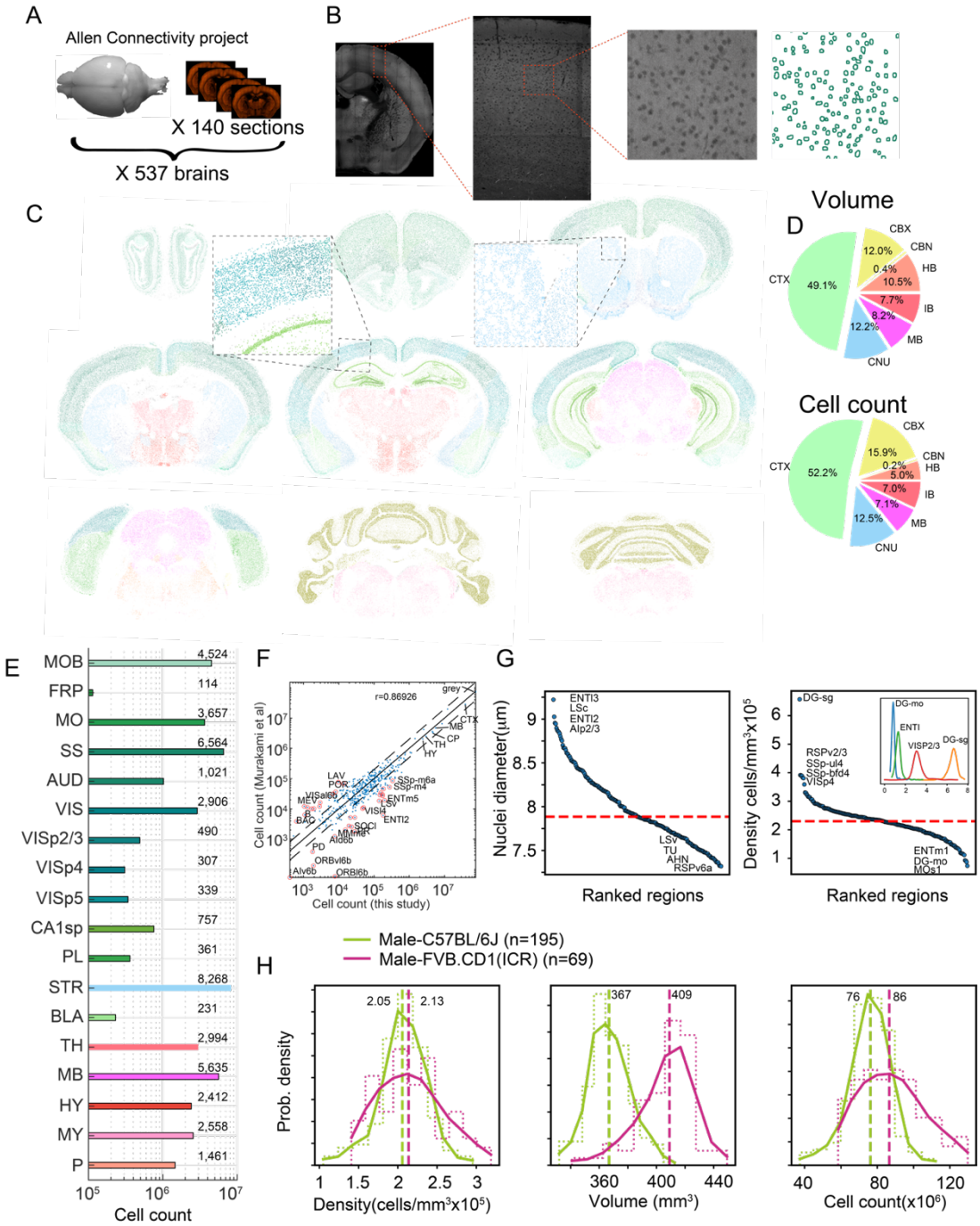

Figure 4: Survey of neuroanatomic properties of the mouse brain.

(A) The analysis is based on a cohort of 537 mouse brains imaged by serial two-photon tomography using the Allen Mouse Brain Connectivity Project (AMBCP). Each brain comprises ~140 coronal sections spaced 100μm apart along the anterior-posterior axis. (B) Example of nucleus segmentation in the isocortex. Each section was divided into tiles of 312×312 pixels (109×109 um) (zoom-ins, right). A trained deep neural network cell segmentation model (see Methods) was applied to detect the contours of nuclei for downstream analysis across tiles, sections, and whole brains, as shown. (C) Segmentation of several sections of one particular brain; segmented nuclei are colored using the Allen Mouse Brain Atlas (AMBA) region convention. (D) Pie charts of the median volumes and cell counts across all 537 brains in the main brain regions, colored using AMBA nomenclature. (E) Median cell counts for selected brain regions in C57BL/6J males (number near bars in thousands). (F) Comparison of region cell counts between this study and Murakami et al., over C57BL/6J males; dots above/below the dashed lines represent regions with greater than two-fold difference. (G) Ranking of 532 regions by nucleus diameter (left) and density (right). Each dot corresponds to the median value of one region over 537 brains. Red dashed line, median across regions. Inset shows distributions of density over 537 brains for selected regions. (H) Distribution of cell density (left), brain volume (middle), and cell count (right), comparing C57BL/6J males and FVB.CD1 males across basic cell groups and regions ("grey").

A resource for exploring neuroanatomical features across regions and populations

To test the power of our model, we explored the densities and nucleus diameter of cortical regions (Figure 5). First, we considered the hippocampal formation (HPF) because imaging-based quantification of its denser regions (pyramidal layers of Ammon's horn and the granule layer of the dentate gyrus) has been difficult[10] and was achieved only recently.[25,28] Analyzing 195 C56BL/6J male brains, we found that the pyramidal layer of CA1 was denser than that of CA3 and CA2, whereas nucleus size was larger in CA3. In the dentate gyrus, the granule layer had the highest density of all regions, with >6.5x10^5 cells/mm^3, and nuclei were largest in the polymorph layer (Figure 5 upper panels). In the isocortex, we examined the extent to which the cortical layers across cortical divisions differed in density and size (Figure 5 lower panels). Layer 1 was consistently underpopulated, having a density of about 10^5 cells/mm^3. The overall rank order from densest to sparsest was maintained, with layer4>layer6a>layer2/3,layer5>layer1, suggesting a similarity in cytoarchitecture between cortical regions. Layer 4 of the primary visual and somatosensory cortices had higher density than did the auditory and visceral cortices. Nucleus diameters showed less distinct distributions between layers, although layer 2/3 and layer 5 tended to have larger nuclei than did layers 4 and 6a.

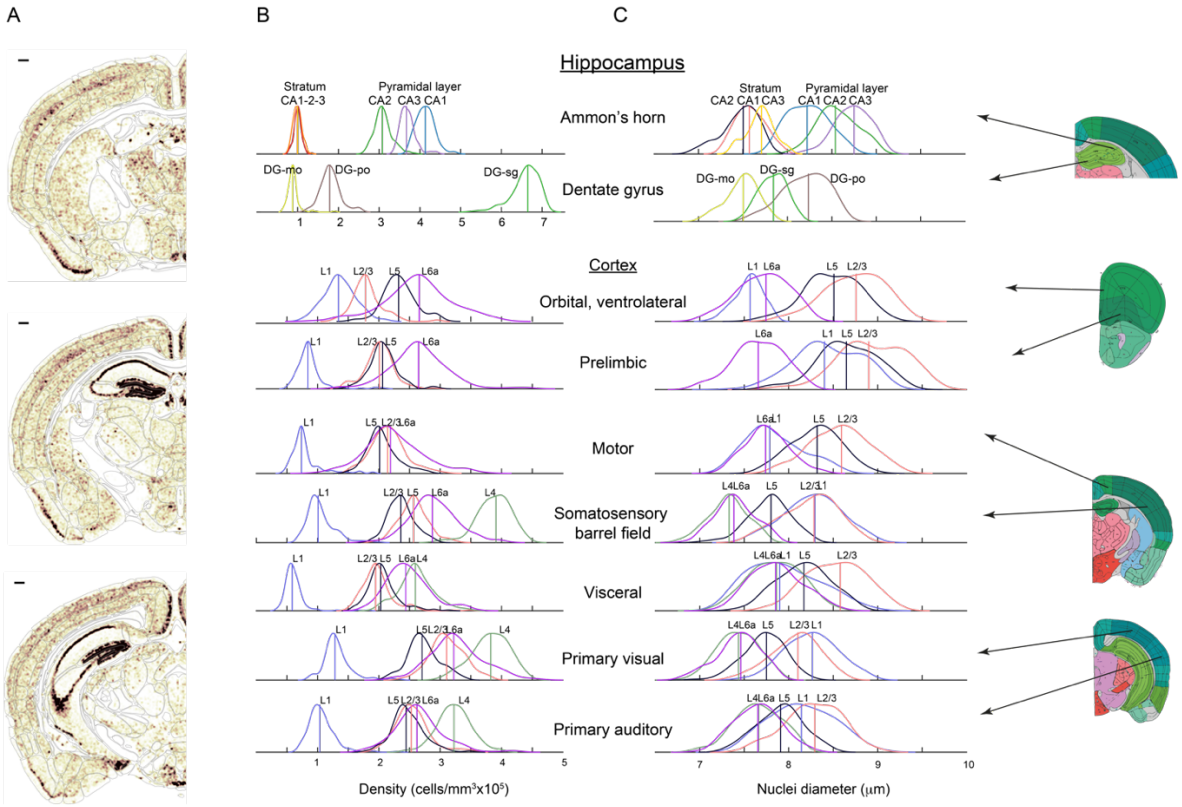

Figure 5. Density and nucleus diameter along cortical regions.

(A) Local density is shown as a heat map over the anatomy of three coronal sections of one brain. White, low; dark brown, high local density; scale bars on upper left corners equal 280μm. Distribution of cell density (B) and nucleus diameter (C) in the hippocampus and selected cortical regions, in 195 C56BL/6J male mice. The two upper rows show Ammon's horn and the dentate gyrus of the hippocampal formation, and the rows below show examples of cortical regions, each resolved to its cortical layers. On the right, approximate locations of each region are indicated in coronal sections of the AMBA.

Regions with volume/density sexual dimorphism in C57BL/6J mice

To examine whether differences in overall brain volume or density (Figure 4H) are isotropic, we analyzed volume, density, and cell count region-specifically. Differences between males and females in regional neuroanatomy have been extensively described, including dimorphic volume and cell count in the medial amygdala (MEA)[29,30] and in the bed nuclei of the stria terminalis (BST).[31] We first compared C57BL/6J males (n=140) with females (n=152). At the global level ("grey"), males and females had similar total numbers of cells (77 and 75 $\times$ 10^6, respectively). These similar counts were achieved differently, however: females had a larger median grey matter volume, whereas males had higher median grey matter density (Figure 6A). We conducted rank sum testing on each region that passed QC (see Methods) for sex differences, in volume and density (Figure 6B). With the notable exception of both MEA and BST, most regions were consistent with the overall trend of larger volumes in females; many were 5-10% larger. Volume sex differences were compensated by higher cell density in the male brains, leading to slightly more cells in most brain regions in males (see also Figure S3A, which shows similar volcano plots for FVB.CD1 mice). We further demonstrated this discordance between median sexual difference in volume vs. density in Fig. 3C, where most brain regions fell in quadrant IV of the volume-density plane. Notable exceptions included the MEA and BST, which were consistently larger in males, and the orbital area layer 2/3, consistently larger in females. Next, we looked beyond the rank sum statistical test, governed by the median of the distribution, at examples of how distributions differ. For example, the ventrolateral orbital area layer 2/3 (ORBvl2/3) showed both larger volumes and slightly higher density in females (Figure 6D left), resulting in significantly more cells in females (Supp. Figure S3A). The opposite was the case for BST, where males had both larger volume and higher density (Figure 6D middle). As a third example, we showed the case of primary auditory area layer 5 (AUDp5), which displayed no difference in region volume, yet density in the male brains was higher (Figure 6D right).

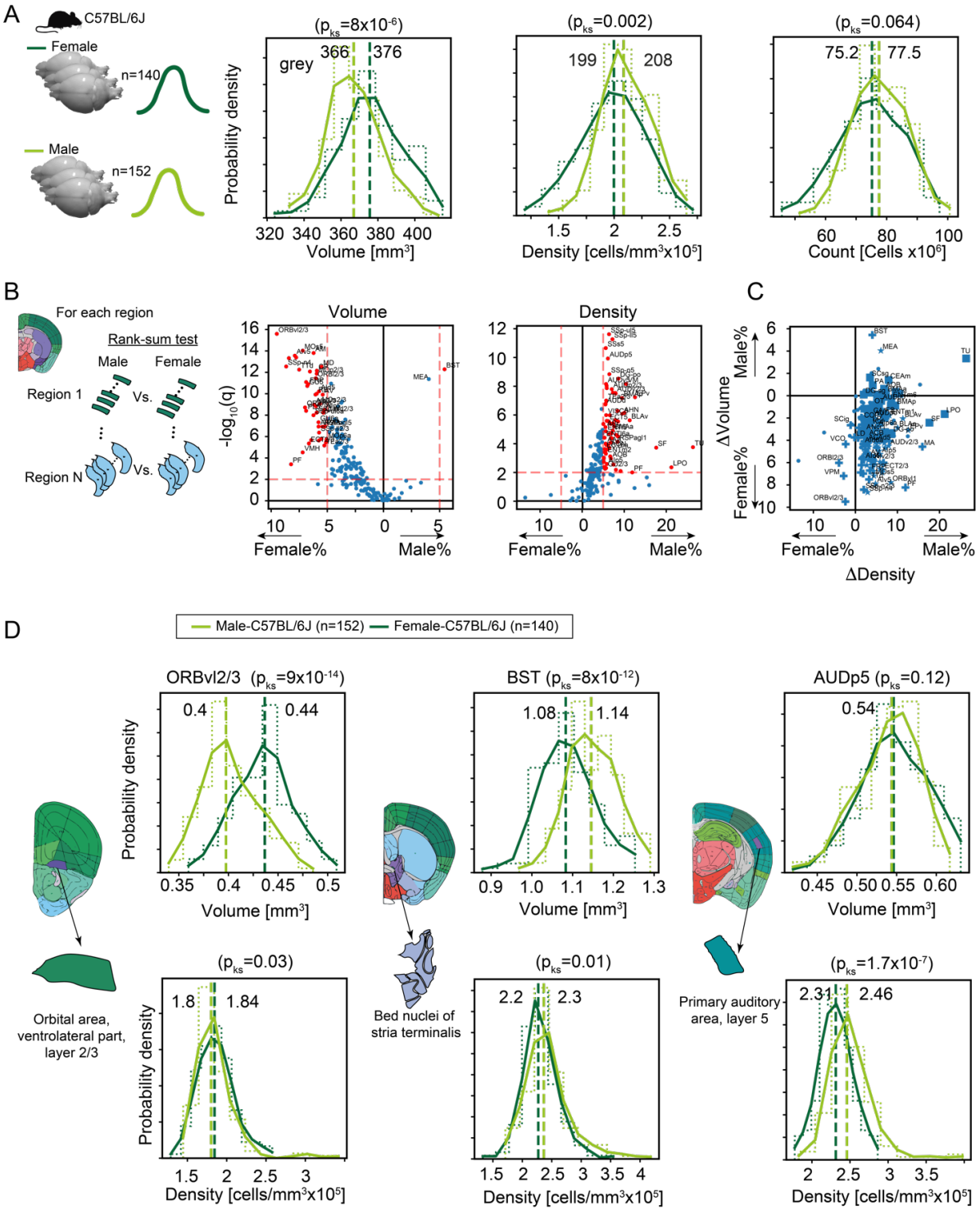

Figure 6. **Sexual dimorphism in C57BL/6J.**

(A) Distribution of volume (left), density (middle), and cell count (right) for the whole brain grey matter ("grey") in female (dark green) and male (light green). P-values correspond to a Kolmogorov-Smirnov test. (B) Volcano plots showing per-region statistical testing for male versus female difference in volume (left) and density (right), each dot representing one region. Horizontal axis, median differences (%); vertical axis, q-values (FDR corrected rank-sum p-values by BH procedure in -log10 scale). Red dots highlight regions with an effect size larger than 5% and q<0.01. (C) Scatter plot of tested regions (dots), showing median differences in volume vs. density. Markers represent statistical significance: both volume and density (star), volume only (+), or density only (square). (D) Examples of regions that display sexually dimorphic volume and/or density. Distributions of volumes appear in the upper row, distributions of densities in the lower row.

Strain differences in volume and density

Following the observation in C57BL/6J mice that female brain volumes were higher despite a smaller body size, we investigated the relation between recorded body weight and grey matter volume. To this end, we added the cohort of outbred FVB.CD1 mice, a strain with 40-50% higher body weight than C57BL/6J. As expected, in both strains, males and females showed distinct distributions for body weight, and males were larger than females (Figure 7A). Distributions for grey matter volume had higher overlap between sexes and showed opposing trends between the strains: in contrast to C57BL/6J, FVB.CD1 females had smaller brain volumes than males. Moreover, within each strain, body weight did not correlate with grey volume. We next quantified sex and strain differences in brain volume and density, resolved to neuroanatomical regions. First, we compared strain differences in females with those in males, showing concordance/discordance patterns between males and females (sex) (Figure 7 B-C and schematic to the right). Second, we compared sex differences in FVB.CD1 with those in C57BL/6J, showing concordance/discordance patterns between strains (Figure 7 D-E and schematic to the right).

Strain-wise analysis: FVB.CD1 brains were overall larger, but the volume expansion with respect to C57BL/6J was not uniform across regions. Region volumes ranged up to 30%, with the extreme example of the cerebellum (CENT2), whose size increased by 50% in both FVB.CD1 males and females (Figure 7B). Moreover, per-region volume differences between strains were, in general, larger in males (i.e., most data points in Figure 7B quadrant III are above the diagonal). Only two regions showed larger volumes in C57BL/6J: the main olfactory bulb (MOB) and the caudal lateral septal nucleus (LSc) (Figure 7B quadrant I).

A similar comparison for cell density per region suggests non-uniform density differences, with almost half the regions being denser in C57BL/6J, and the other half in FVB.CD1 (Figure 7B quadrants I and III, respectively). In this comparison, olfaction-related regions (AOB and MOB) showed higher density in C57BL/6J, while the LSc showed the opposite effect.

Sex-wise analysis: Differences in volume confirmed sexual dimorphism in MEA and BST, which were larger in males for both strains. These differences were more pronounced in FVB.CD1 than in C57BL/6J (Figure 7D quadrant I). Many brain regions showed “strain-discordant” dimorphism, with females having a larger volume in C57BL/6J and males in FVB.CD1 (Figure 6D quadrant II). Although total brain volume in FVB.CD1 males was larger, some regions showed larger volume in females (e.g., the previously mentioned orbital cortex ORB, Figure 6D quadrant III). Comparing sexual dimorphism in density (Figure 6E), we found a simpler and more consistent picture: in both strains, males had higher density in all regions except for ORBvl2/3. Note that in density as well, sex differences were found to be larger in FVB.CD1 (most data points in Figure 6E quadrant I are above the diagonal).

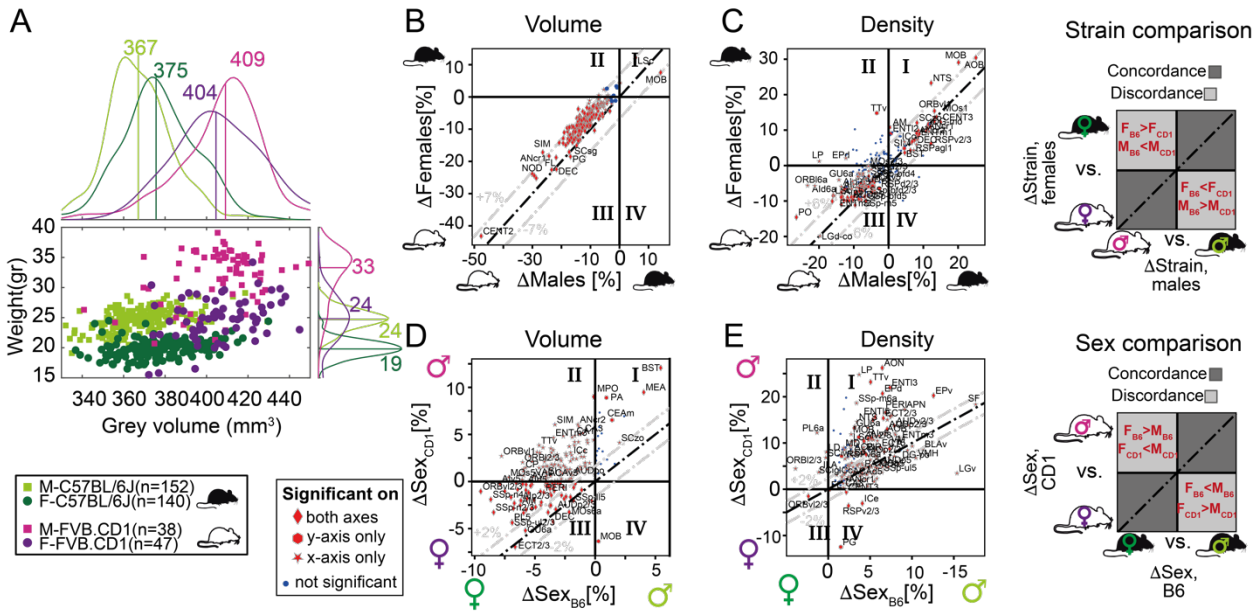

Figure 7. **Sexual and cross-strain dimorphism in C57BL/6J (B6) and FVB.CD1 (CD1).**

(A) Scatter plot showing body weight vs. grey volume for 537 brains. Side panels show the group distributions of grey matter volume (upper) and weight (right). Lines are the medians whose values are indicated. (B-C) Strain comparison of per region volume (B) and density (C). Differences between the median values of the strains, per region, are shown for males (horizontal axis) and females (vertical axis). Points in quadrants I and III suggest concordance between males and females across strains, as illustrated in the schematic on the right. Red markers designate statistical significance in either axes or in both. (D-E) Sex comparison of per region volume (D) and density (E). Points in quadrants I and III suggest concordance between C57BL/6J and FVB.CD1 across sex, as illustrated in the schematic on the right.

Region-wise correlations between volume and density across brains

To the best of our knowledge, no previous study simultaneously quantified cell density (*D*) and brain region volume (*V*). We therefore sought to investigate whether constraints exist between *D* and *V*. For example, if the number of cells in a region is constant across brains, *D* and *V* must be negatively correlated. If, by contrast, the number of cells in a region, *N*, scales with the volume while *D* remains constant, *D* and *V* display zero correlation. If a positive correlation exists between *D* and *V*, the number of cells *N* grows faster than linear with respect to either *D* or *V* (Figure 8A). Based on per-region measurements of both *V* and *D*, we calculated regionally-resolved Pearson correlations between volume and density (Figure 8B). In 72% of regions (289/397), cell density was negatively correlated with volume (Figure 8C), with a median correlation of -0.096. For example, we showed two regions where *N* was positively correlated with both *D* and *V*, yet the correlation between *D* and *V* was either positive (AAA) or negative (SSs2/3). This suggests that for some regions, cell count does not scale simply or linearly with volume.

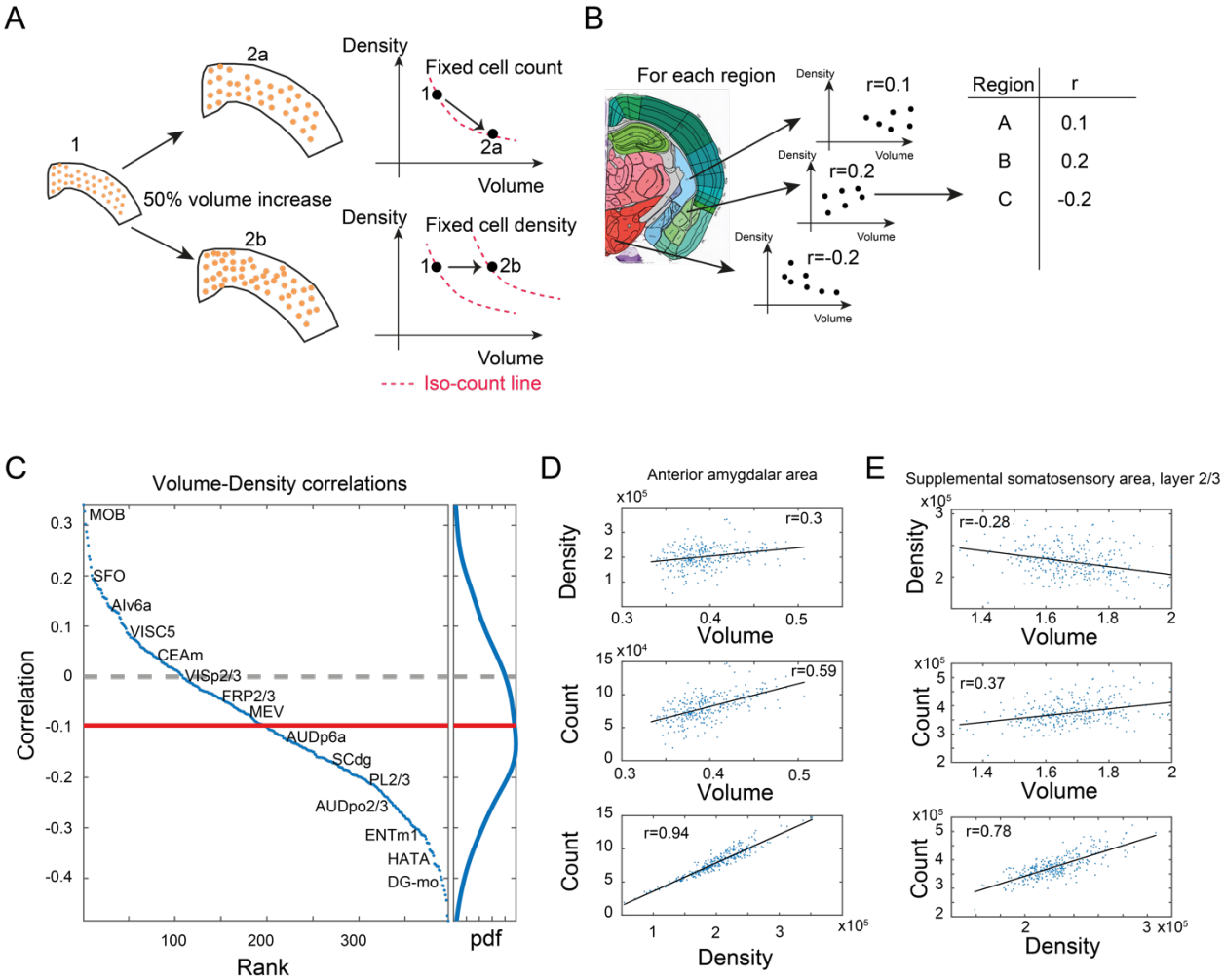

Figure 8. **Correlations between volume, cell count, and density.**

(A) Schematic illustration of two types of relations between regional cell density and volume, associating region expansion with a fixed number of cells (upper) or with a fixed density (lower). Each regional expansion can be represented by a shift in the volume-density plane (right column). (B) A scheme showing how for each region the correlation between density and volume was measured over the whole dataset. (C) Brain regions ranked by the correlation between volume and density. Correlations higher than 0.13 or lower than -0.13 correspond to q-values lower than 0.05. Side panel displays the distribution of correlation values, and its median is denoted by the red line. (D) Correlations between volume, density, and count in the anterior amygdalar area. (E) As (D), for supplemental somatosensory area, layer 2/3.

Inter-brain similarity between regions based on volume and density

Finally, we assessed similarity between regions, based on volume or density. We used tSNE as a 2D embedding method over the density data (Figure 9A-B). Briefly, each region is characterized by a vector of 537 components, each representing its density across one brain. 2D embedding aims to preserve the local similarity between regions. The tSNE embedding map in Figure 9A reveals clear 2D "clusters," largely consistent with neuroanatomical classification. Cortical regions appear in the upper part of the map (colored green), and cerebellum (yellow), midbrain, and hindbrain in the lower part. We further explored whether the order within the cortical part may be explained by layer structure or by cortical division, but found no clear structure (Figure S4 A-B). Compared to Figure 9A, tSNE embedding based on volume was more "dispersed" and displayed disorder with respect to neuroanatomical classification (Figure 9C-D). To demonstrate

that the tSNE map is indeed based on true variations in region-to-region correlations, we compared density-based with volume-based correlations. First, for each region we identified its 10 most correlated regions based on either density or volume. These correlation values were higher for density-based correlations across almost all regions (Figure S4C), supporting the observed density-based "order." Second, we selected 10 representative regions across the brain and calculated all their pairwise correlations (Figure S4D), showing that even for distant regions, density-based correlations remain much higher than volume-based correlations. Thus, similarity in volume across regions is less "preserved" brain-wide than similarity in cell density.

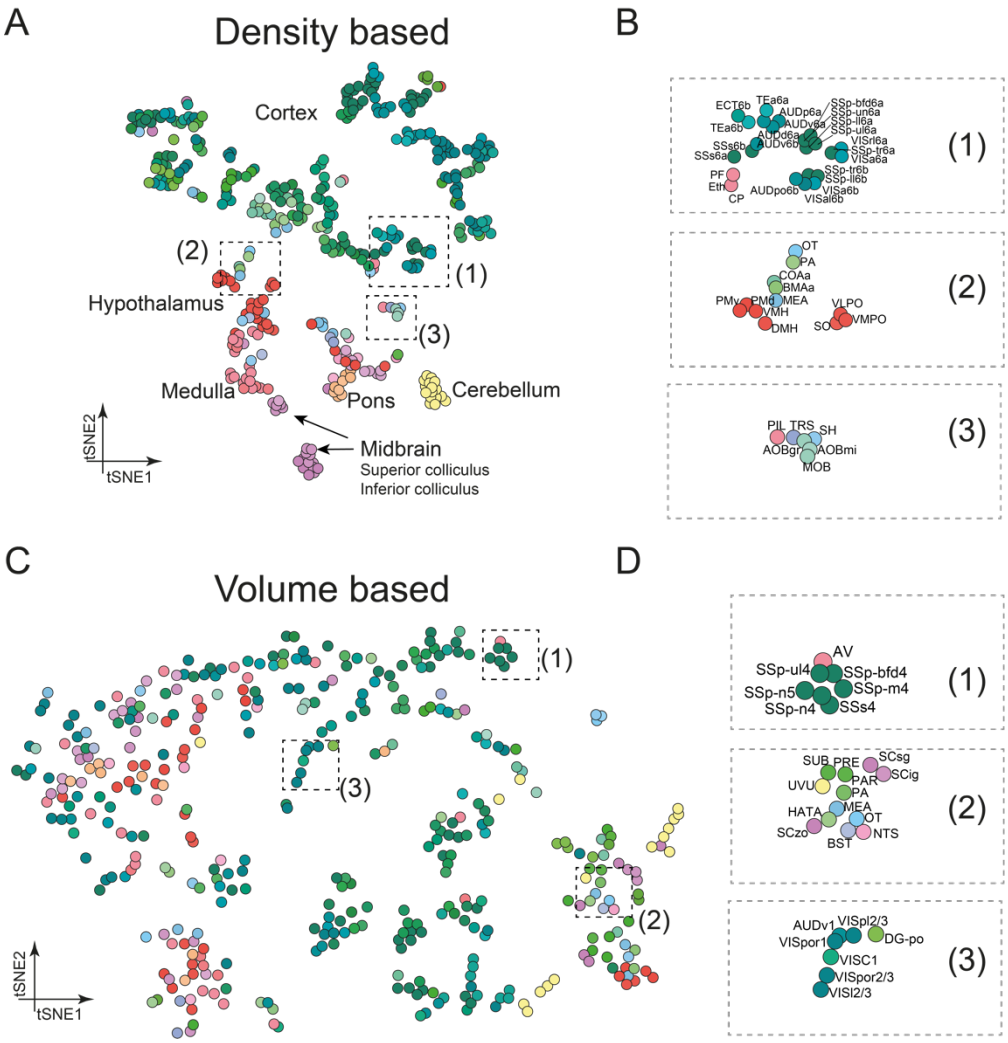

Figure 9. **Visualizing the similarity between brain regions based on ABCP.**

A tSNE embedding of brain regions based on pairwise correlations between region density (A) or volume (C). Each dot represents a region and is colored according to the AMBA convention. (B,D) Zoom-in on three frames from (A) and (C), respectively.

Discussion

We presented an automated, imaging-based, staining-free study of neuroanatomy and cytoarchitecture in the mouse brain. We conducted our measurements on a massive, high-quality dataset of serial two-photon tomography,[13] aligned with a well-annotated reference atlas.[19] This made possible, for the first time, a detailed population-wide analysis of two important neuroanatomical variables simultaneously: cell density and volume, resolved for 532 regions. The data spans an unprecedented cohort of 537 mice of two strains, the inbred C57BL/6J, and the hybrid FVB1.CD-1, each represented by both females and males. Our high-throughput measurements of cell densities were achieved by using a DNN trained to detect low-autofluorescent cell nuclei with high accuracy, even in the most cell-dense regions of the brain.

The study has several limitations. First, the model is sensitive to image quality, and in particular, contrast between dark nuclei and autofluorescent surroundings. In the hindbrain (pons, medulla), contrast was exceedingly weak, and we expect our quantifications in this region to strongly underestimate real cell densities, to an extent we cannot quantify. Second, AMBA annotations were not always resolved to the most refined level of the atlas hierarchy. For example, density values for the cerebellum appear to be uncharacteristic because the cell-dense granule layer and sparse molecular layer were not distinguished at the deepest level of annotation (e.g., CENT3 included the granular and molecular layers). The same is true for the hippocampus CA1-2-3, where we used cell density-based clustering to distinguish the pyramidal layer (sp) from its surrounding sparse layers (slm, so, sr, see Methods). Therefore, although the model performed exceedingly well even in these cell-dense regions, the absence of annotations stood occasionally in the way of making biologically meaningful distinctions.

Nevertheless, we provided key statistics that help answer fundamental, recurring questions in neuroanatomy. Although no other study presented simultaneous measurements of volume and cell density, our data correlate well with a wealth of literature in the field. We achieved good region-wise correlation with full 3D volumetric cell counts by expansion microscopy[25] (Fig 1F). Our derived cell count of mouse brain grey matter (76×10^6 for male C57BL/6J) is well within the range of existing cell count estimates for adult males ($67 - 150 \times 10^6$ cells),[32,33,25,28]

By measuring the largest cohort to date, we provided partial support for the notion that this extreme range in the literature may not stem from variation in strain or sex, but rather from individual differences.[25] The median cell counts between sexes and strains differed no more than 13% overall, or ~40% for the most deviant individual structures (MOB and CENT2), while the standard deviation across individuals was ~10 million cells (for C57BL/6J male alone). Hence, values between $55 \times 10^6 - 95 \times 10^6$ are within ±2 of the distribution of total grey matter cell counts. We claim that the notion of "ground truth" values of brain cell number *can* be reached, yet are best reflected by a population distribution. Our dataset provides a large, important corpus toward this "ground truth," and similar studies can further help distinguish technical biases from true biological variation.

We validated the existing literature describing examples of regions that show sex- or strain-based differences, and expanded on it. For instance, medial amygdala and bed nuclei of the stria terminalis were both larger and denser in males, but to a lesser extent than reported in smaller studies[34] and to a similar extent to what was reported in MRI-based studies.[35] By contrast, in

females, several prefrontal cortex structures were larger (e.g., ORBvl2/3), which resulted in higher cell counts. We found no evidence of this phenomenon in the literature on mice, but an MRI population study of 2,838 human individuals found higher grey matter volume (GMV) in prefrontal areas in women.[36] Between the strains, we found considerable differences in the olfactory system, which was larger and denser in C57BL/6J, and in the cerebellum, which was larger in FVB.CD-1. Finally, we provide an accessible, web-based platform for open exploration of the data. The web application allows researchers to freely compute distributions of any measured neuroanatomical features, for any brain region, and across the entire population or specific subsets. This exploratory resource can be of great use for experimental design, and lead to more accurate brain modeling.

Bibliography

1. Rumelhart, D., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* (1986) doi:10.1038/323533a0.
2. O'Shea, K. & Nash, R. An Introduction to Convolutional Neural Networks. *arXiv:1511.08458 [cs]* (2015).
3. He, K., Zhang, X., Ren, S. & Sun, J. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]* (2015).
4. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. in *2009 IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (2009). doi:10.1109/CVPR.2009.5206848.
5. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. *arXiv:1703.06870 [cs]* (2018).
6. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick, R. Detectron2. (2019).
7. How_MaskRCNN_works | ArcGIS Developer. https://developers.arcgis.com/python/guide/how-maskrcnn-works/.
8. White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode Caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci* **314**, 1–340 (1986).
9. Keller, D., Erö, C. & Markram, H. Cell Densities in the Mouse Brain: A Systematic Review. *Frontiers in Neuroanatomy* **12**, (2018).
10. Attili, S. M., Silva, M. F. M., Nguyen, T. & Ascoli, G. A. Cell numbers, distribution, shape, and regional variation throughout the murine hippocampal formation from the adult brain Allen Reference Atlas. *Brain Struct Funct* **224**, 2883–2897 (2019).
11. Ueda, H. R. *et al.* Whole-Brain Profiling of Cells and Circuits in Mammals by Tissue Clearing and Light-Sheet Microscopy. *Neuron* **106**, 369–387 (2020).
12. Ragan, T. *et al.* Serial two-photon tomography for automated ex vivo mouse brain imaging. *Nat Methods* **9**, 255–258 (2012).
13. Oh, S. W. *et al.* A mesoscale connectome of the mouse brain. *Nature* **508**, 207–214 (2014).
14. [1512.03385v1] Deep Residual Learning for Image Recognition. https://arxiv.org/abs/1512.03385v1.
15. Dutta, A. & Zisserman, A. The VGG Image Annotator (VIA). *CoRR* **abs/1904.10699**, (2019).
16. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. in *Advances in Neural Information Processing Systems* vol. 25 (Curran Associates, Inc., 2012).
17. [1412.6980] Adam: A Method for Stochastic Optimization. https://arxiv.org/abs/1412.6980.
18. Jaccard, P. The Distribution of the Flora in the Alpine Zone.1. *New Phytologist* **11**, 37–50 (1912).
19. Wang, Q. *et al.* The Allen Mouse Brain Common Coordinate Framework: A 3D Reference Atlas. *Cell* **181**, 936-953.e20 (2020).
20. Amato, S. P., Pan, F., Schwartz, J. & Ragan, T. M. Whole Brain Imaging with Serial Two-Photon Tomography. *Frontiers in Neuroanatomy* **10**, (2016).
21. Allen Mouse Brain Connectivity Atlas TECHNICAL WHITE PAPER : OVERVIEW OVERVIEW The Allen Mouse Brain. in (2016).
22. Costantini, I. *et al.* Autofluorescence enhancement for label-free imaging of myelinated fibers in mammalian brains. *Scientific Reports* **11**, 8038 (2021).

23. DaCosta, R. S., Andersson, H., Cirocco, M., Marcon, N. E. & Wilson, B. C. Autofluorescence characterisation of isolated whole crypts and primary cultured human epithelial cells from normal, hyperplastic, and adenomatous colonic mucosa. *J Clin Pathol* **58**, 766 (2005).
24. Kretschmer, S. *et al.* Autofluorescence multiphoton microscopy for visualization of tissue morphology and cellular dynamics in murine and human airways. *Lab Invest* **96**, 918–931 (2016).
25. Murakami, T. C. *et al.* A three-dimensional single-cell-resolution whole-brain atlas using CUBIC-X expansion microscopy and tissue clearing. *Nature Neuroscience* **21**, 625–637 (2018).
26. Neumann, F. R. & Nurse, P. Nuclear size control in fission yeast. *J Cell Biol* **179**, 593–600 (2007).
27. Huber, M. D. & Gerace, L. The size-wise nucleus: nuclear volume control in eukaryotes. *Journal of Cell Biology* **179**, 583–584 (2007).
28. Seiriki, K. *et al.* High-Speed and Scalable Whole-Brain Imaging in Rodents and Primates. *Neuron* **94**, 1085-1100.e6 (2017).
29. Morris, J. A., Jordan, C. L. & Breedlove, S. M. Sexual dimorphism in neuronal number of the posterodorsal medial amygdala is independent of circulating androgens and regional volume in adult rats. *Journal of Comparative Neurology* **506**, 851–859 (2008).
30. Morris, J. A., Jordan, C. L., Dugger, B. N. & Breedlove, S. M. Partial demasculinization of several brain regions in adult male (XY) rats with a dysfunctional androgen receptor gene. *Journal of Comparative Neurology* **487**, 217–226 (2005).
31. Garcia-Falgueras, A. *et al.* The expression of brain sexual dimorphism in artificial selection of rat strains. *Brain Research* **1052**, 130–138 (2005).
32. Herculano-Houzel, S. *et al.* Updated Neuronal Scaling Rules for the Brains of Glires (Rodents/Lagomorphs). *BBE* **78**, 302–314 (2011).
33. Herculano-Houzel, S., Catania, K., Manger, P. R. & Kaas, J. H. Mammalian Brains Are Made of These: A Dataset of the Numbers and Densities of Neuronal and Nonneuronal Cells in the Brain of Glires, Primates, Scandentia, Eulipotyphlans, Afrotherians and Artiodactyls, and Their Relationship with Body Mass. *BBE* **86**, 145–163 (2015).
34. Cooke, B. M., Stokas, M. R. & Woolley, C. S. Morphological sex differences and laterality in the prepubertal medial amygdala. *J Comp Neurol* **501**, 904–915 (2007).
35. Qiu, L. R. *et al.* Mouse MRI shows brain areas relatively larger in males emerge before those larger in females. *Nat Commun* **9**, 2615 (2018).
36. Lotze, M. *et al.* Novel findings from 2,838 Adult Brains on Sex Differences in Gray Matter Brain Volume. *Sci Rep* **9**, 1671 (2019).

Supplementary Material

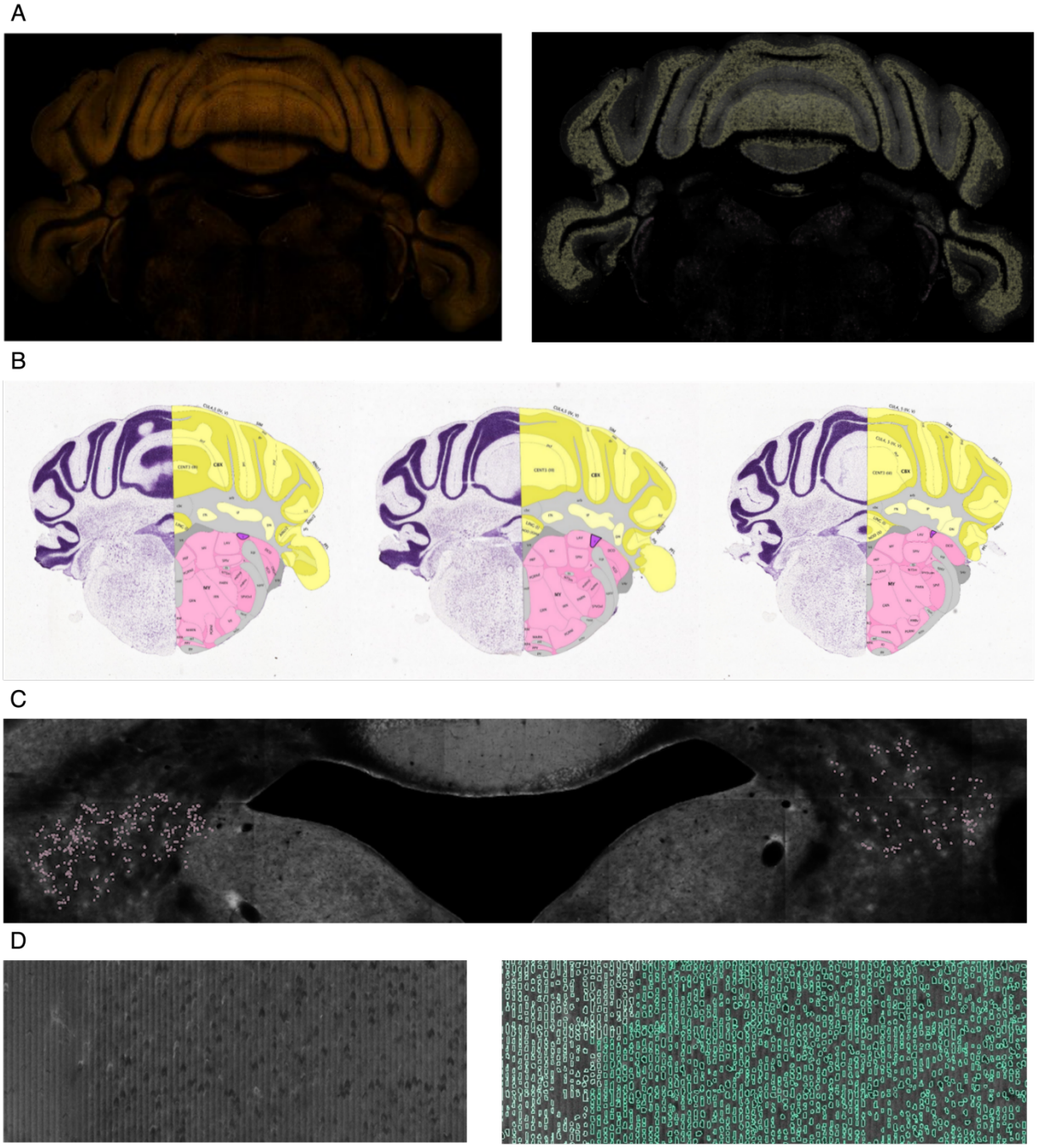

Figure S1. Sample brains and regions filtered out.

(A) Brains displaying dark images. Brightness level is too low, especially in the hindbrain region. Experiment #112672268 section 110. (B) Small region sample. Sections 115, 116 and 117 from the Allen Brain Atlas with 'Nucleus y' region highlighted in violet on the right hemisphere. (C) Region displaying different estimates in right vs. left hemispheres. Clearly much more cells detected in the left hemisphere of the 'Superior vestibular nucleus' region due to brightness differences. Experiment #113933871 section 113. (D) Brain displaying outliers in cell count. The model fails to produce accurate predictions and detects too many cells due to the noise. Experiment #268399145 section 77.

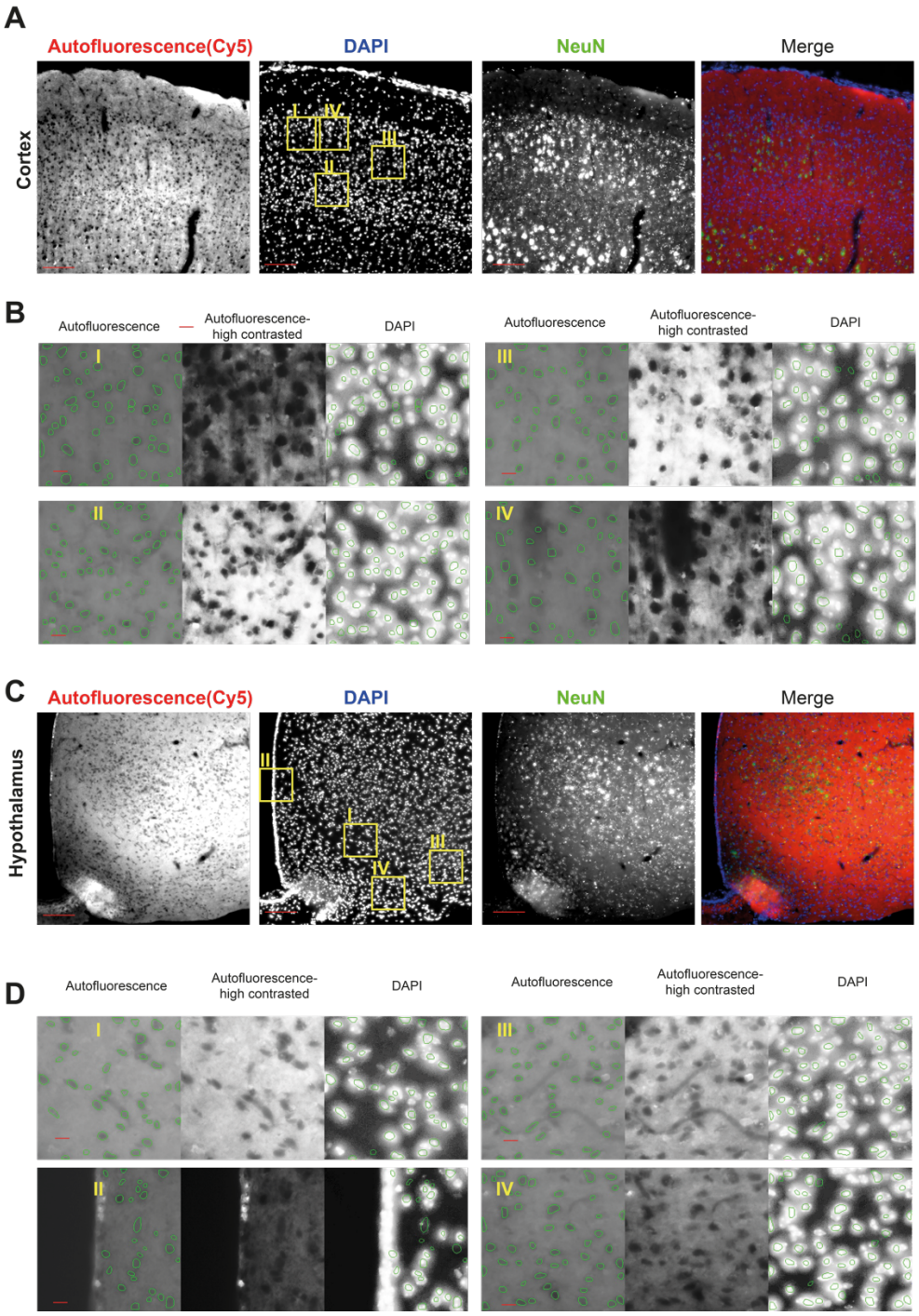

Figure S2 (related to Figure 4): Autofluorescence signal corresponds to nucleus validation.

A mouse brain was perfused with 4% PFA followed by sectioning and DAPI staining. Representative image from cortex **(A)** or hypothalamus **(C)** showing autofluorescence (cy5-far red), DAPI, NeuN, and the merge of the three channels. **(B,D)** We applied the same segmentation DNN used for the Allen Mouse Connectivity dataset. Each tile in **(B)** and **(D)** shows detected objects on top of the original images (left), autofluorescence high contrast (middle), and DAPI overlayed with the same objects (right).

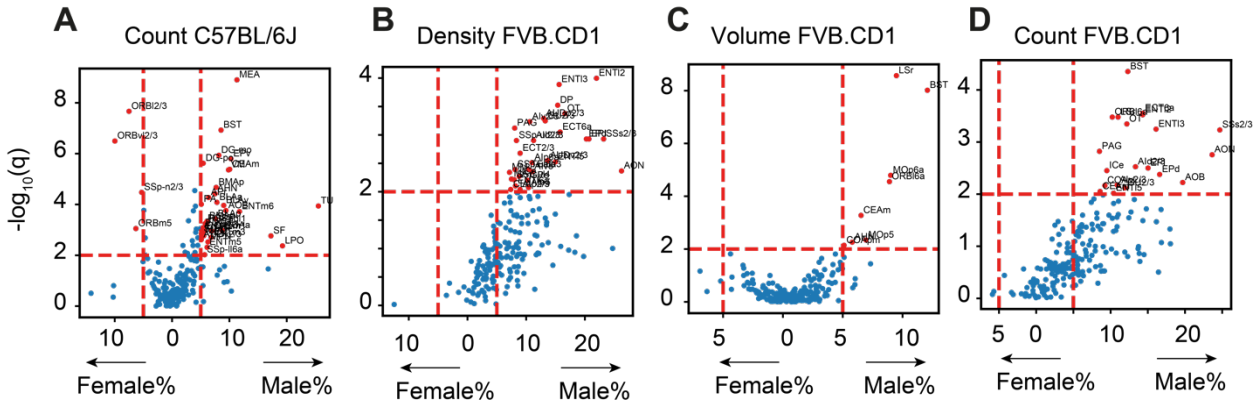

Figure S3 (related to Figure 5): Volcano plots showing per-region statistical testing for male vs. female differences.
The horizontal axis represents median differences (%) and the vertical axis displays the q-values (FDR corrected rank-sum p-values by BH procedure in -log10 scale). Red dots correspond to an effect size larger than 5% and q<0.01. **(A)** Cell count in C57BL/6J, **(B)** Density in FVB.CD1, **(C)** Volume in FVB.CD1, **(D)** Cell count in FVB.CD1.

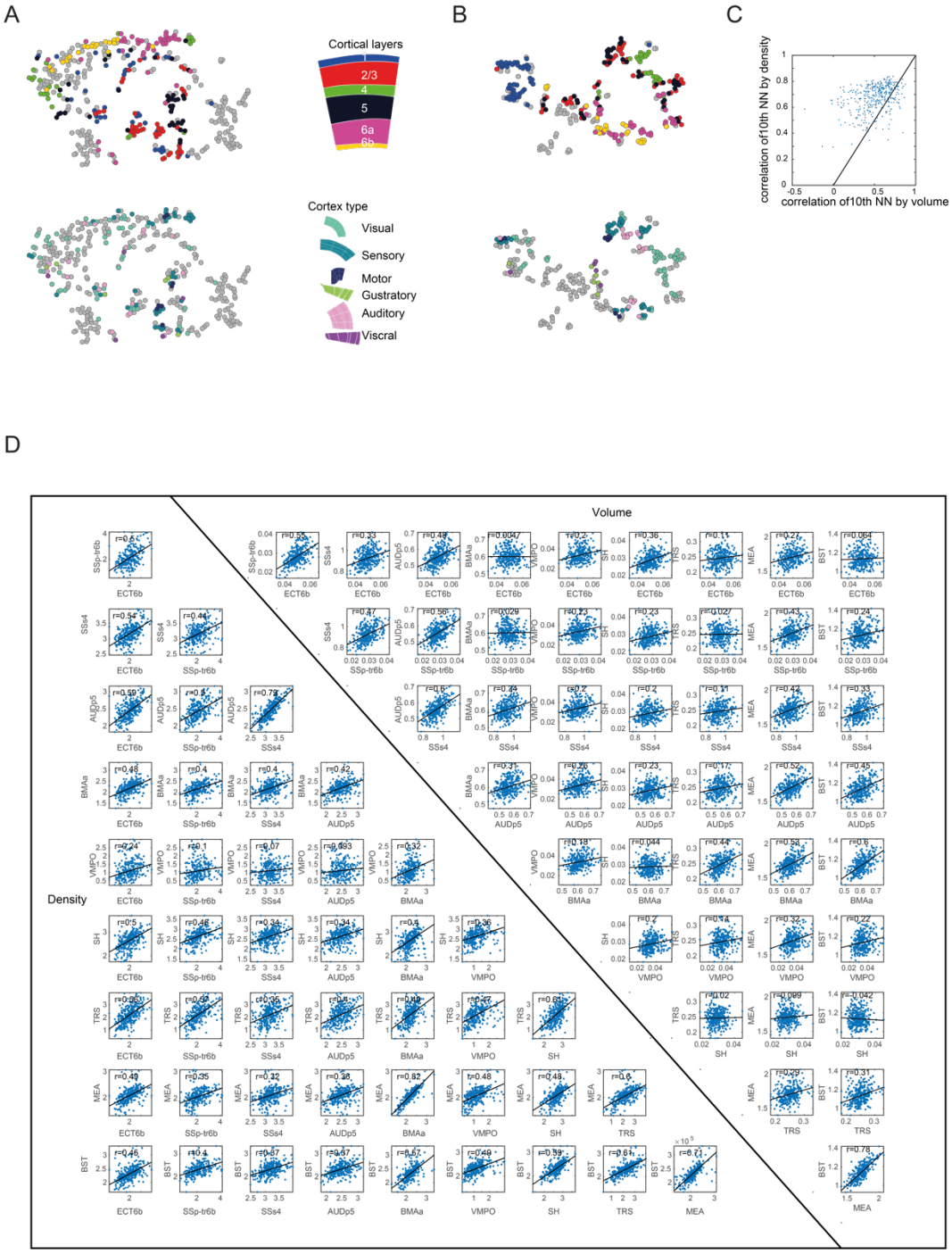

Figure S4 (related to Figure 9):
(A) The density based tSNE plot of Figure 9A color-labelled according to cortical layers (upper) and cortical division (lower). **(B)** The same for the volume based tSNE plot of Figure 9C. **(C)** Correlation to the 10th nearest neighbor for each region when using volume (horizontal axis) or density (vertical axis). **(D)** Examples for region-region correlations. We show regions ECT6b, SSp-tr6b, SSs4, AUDp5, BMAa, VMPO, SH, TRS, MEA, and BST. Correlations are calculated by density (lower triangle) and volume (upper triangle).

Table S1: List of the processed brain regions with the corresponding macroscopic properties

AMBA brain region abbreviation	Volume, mm3	Density, cell/mm3	Cell count	Cell diameter, um	Is Leaf?
AAA	0.39	198760	79212	7.56	TRUE
ACA	4.58	213077	974059	8.27	FALSE
ACAd	2.58	199596	515775	8.34	FALSE
ACAd1	0.51	86452	44455	8.19	TRUE
ACAd2/3	0.50	212890	107693	8.73	TRUE
ACAd5	0.95	216193	205904	8.45	TRUE
ACAd6a	0.61	260192	154335	7.60	TRUE
ACAd6b	0.02	132124	2868	7.06	TRUE
ACAv	1.99	228593	453959	8.20	FALSE
ACAv1	0.40	91238	36259	7.76	TRUE
ACAv2/3	0.42	238335	101638	8.52	TRUE
ACAv5	0.82	260186	214167	8.26	TRUE
ACAv6a	0.29	306983	88937	7.62	TRUE
ACAv6b	0.05	224306	11330	7.18	TRUE
ACB	3.54	306270	1093567	7.70	TRUE
AD	0.14	220954	30133	8.39	TRUE
ADP	0.08	272235	20699	7.29	TRUE
AHN	0.56	238982	136628	7.34	TRUE
AI	6.48	154582	1006649	8.47	FALSE
AId	3.04	168296	509153	8.43	FALSE
AId1	0.41	46317	19111	7.50	TRUE
AId2/3	0.72	147499	105880	8.81	TRUE
AId5	1.20	183640	221739	8.48	TRUE
AId6a	0.66	232249	151684	8.02	TRUE
AId6b	0.04	198938	8048	7.87	TRUE
AIp	2.01	141619	284403	8.52	FALSE
AIp1	0.38	53760	20004	7.13	TRUE
AIp2/3	0.65	147253	95624	8.87	TRUE
AIp5	0.61	152637	93179	8.55	TRUE
AIp6a	0.35	202016	69526	7.88	TRUE
AIp6b	0.02	205517	3601	7.42	TRUE
AIv	1.46	145445	213027	8.46	FALSE
AIv1	0.22	57748	12737	7.55	TRUE
AIv2/3	0.50	134971	67932	8.74	TRUE
AIv5	0.58	173871	101536	8.36	TRUE
AIv6a	0.15	199923	29543	8.08	TRUE
AIv6b	0.00	199895	448	8.07	TRUE
AM	0.32	186332	60184	8.20	FALSE
AMd	0.19	180279	33541	8.31	TRUE

AMv	0.14	193355	26435	8.05	TRUE
AN	9.47	270120	2575683	7.91	FALSE
ANcr1	4.89	275829	1364171	7.88	TRUE
ANcr2	4.57	263512	1210129	7.95	TRUE
AOB	0.56	236602	131693	7.65	FALSE
AOBgl	0.14	188441	26264	7.39	TRUE
AOBgr	0.18	274137	50805	7.82	TRUE
AOBmi	0.23	233655	54332	7.59	TRUE
AON	3.87	196427	756448	7.73	TRUE
AP	0.04	197773	8298	7.64	TRUE
APN	1.00	214702	214853	7.66	TRUE
APr	0.31	211244	65278	8.16	TRUE
ARH	0.22	151715	33481	7.40	TRUE
ATN	1.73	200215	345224	8.19	FALSE
AUD	4.73	222520	1045656	8.01	FALSE
AUDd	1.00	226468	225216	7.96	FALSE
AUDd1	0.19	93426	17595	8.00	TRUE
AUDd2/3	0.25	243424	61386	8.22	TRUE
AUDd4	0.10	317147	33241	7.64	TRUE
AUDd5	0.27	236213	63681	7.96	TRUE
AUDd6a	0.15	281973	43126	7.69	TRUE
AUDd6b	0.03	212228	5256	7.52	TRUE
AUDp	1.75	224668	392341	7.96	FALSE
AUDp1	0.37	103573	37905	8.07	TRUE
AUDp2/3	0.36	250406	90783	8.24	TRUE
AUDp4	0.19	318734	58332	7.67	TRUE
AUDp5	0.56	241599	134219	7.91	TRUE
AUDp6a	0.24	261604	61080	7.68	TRUE
AUDp6b	0.05	208278	9519	7.48	TRUE
AUDpo	0.50	241860	122076	7.87	FALSE
AUDpo1	0.11	113286	12646	7.91	TRUE
AUDpo2/3	0.11	278043	30981	8.11	TRUE
AUDpo4	0.06	340258	19842	7.58	TRUE
AUDpo5	0.14	256990	36898	7.89	TRUE
AUDpo6a	0.06	289104	18333	7.61	TRUE
AUDpo6b	0.01	191239	2760	7.33	TRUE
AUDv	1.48	208148	304977	8.15	FALSE
AUDv1	0.27	76418	20798	8.10	TRUE
AUDv2/3	0.30	215672	64330	8.48	TRUE
AUDv4	0.16	266198	42649	7.99	TRUE
AUDv5	0.46	232007	107015	8.11	TRUE

AUDv6a	0.24	265004	61821	7.81	TRUE
AUDv6b	0.04	206860	7788	7.64	TRUE
AV	0.35	252970	87755	8.18	TRUE
AVP	0.07	202864	14672	7.30	TRUE
AVPV	0.14	171676	24066	7.29	TRUE
B	0.01	219158	1182	7.55	TRUE
BA	0.02	101156	1760	7.58	TRUE
BAC	0.00	181646	651	6.96	TRUE
BLA	1.50	156164	233583	8.41	FALSE
BLAa	0.61	174659	106244	8.11	TRUE
BLAp	0.56	157008	88714	8.73	TRUE
BLAv	0.33	116694	37815	8.21	TRUE
BMA	1.17	198784	234459	7.99	FALSE
BMAa	0.60	215584	129880	7.55	TRUE
BMAp	0.57	181097	103310	8.38	TRUE
BS	99.10	154733	15173684	7.67	FALSE
BST	1.13	230551	261030	7.44	TRUE
CA	13.98	167380	2359727	8.27	FALSE
CA1	8.23	168577	1400025	8.12	FALSE
CA1sp	1.90	415484	777277	8.31	TRUE
CA1sr	6.33	96850	616075	7.65	TRUE
CA2	0.43	147854	65212	8.22	FALSE
CA2sp	0.10	314352	31503	8.52	TRUE
CA2sr	0.33	99807	33370	7.53	TRUE
CA3	5.34	166231	891194	8.47	FALSE
CA3sp	1.30	376990	485766	8.75	TRUE
CA3sr	4.04	99940	402894	7.74	TRUE
CB	47.25	261237	12534804	7.88	FALSE
CBN	1.47	125844	187222	7.73	FALSE
CBX	45.03	268401	12272554	7.88	FALSE
CEA	1.04	208397	217243	7.72	FALSE
CEAc	0.24	196820	46711	7.98	TRUE
CEAl	0.20	257507	51410	7.89	TRUE
CEAm	0.60	196108	119013	7.45	TRUE
CENT	3.47	289608	1026539	7.85	FALSE
CENT2	1.16	289957	339262	7.88	TRUE
CENT3	2.32	289320	691056	7.83	TRUE
CH	231.33	214293	49375658	8.01	FALSE
CL	0.30	198882	58878	7.86	TRUE
CLA	0.38	236120	89395	8.01	TRUE
CM	0.20	238338	47865	8.00	TRUE

CN	1.36	139454	192093	7.65	FALSE
CNU	45.97	213381	9744094	7.77	FALSE
COA	2.70	153554	416834	8.23	FALSE
COAa	0.61	184154	112494	7.53	TRUE
COAp	2.09	144228	303258	8.38	FALSE
COApl	0.98	132083	128349	8.51	TRUE
COApm	1.12	154956	172650	8.29	TRUE
COPY	2.24	292193	662082	7.86	TRUE
CP	22.19	214194	4719098	7.83	TRUE
CS	0.49	125748	61082	7.38	TRUE
CTX	184.98	213867	39409417	8.06	FALSE
CTXpl	177.98	214885	38163576	8.06	FALSE
CTXsp	7.09	178352	1260008	8.09	FALSE
CU	0.29	151366	43077	7.65	TRUE
CUL	6.00	274568	1660279	7.82	FALSE
CUL4, 5	6.00	274568	1660279	7.82	TRUE
CUN	0.45	129710	58130	7.52	TRUE
DCN	0.37	141575	52814	7.66	FALSE
DCO	0.50	183952	93692	7.83	TRUE
DEC	3.15	243357	780036	7.90	TRUE
DG	5.52	273641	1514574	7.92	FALSE
DG-mo	3.34	83901	281281	7.49	TRUE
DG-po	0.43	174631	74807	8.15	TRUE
DG-sg	1.75	660261	1153040	7.96	TRUE
DMH	0.29	271513	77833	7.48	TRUE
DMX	0.14	184893	26175	8.03	TRUE
DN	0.27	197294	53450	7.71	TRUE
DORpm	10.42	193453	2001049	7.93	FALSE
DORsm	5.49	167878	923931	7.80	FALSE
DP	0.40	169684	67859	8.40	TRUE
DR	0.12	232365	27935	7.65	TRUE
DTN	0.07	346151	25635	7.51	TRUE
ECT	1.42	174210	244405	8.43	FALSE
ECT1	0.26	56654	14993	8.06	TRUE
ECT2/3	0.38	188663	71568	8.65	TRUE
ECT5	0.41	193391	78612	8.47	TRUE
ECT6a	0.31	225834	69288	8.23	TRUE
ECT6b	0.05	195316	9498	7.82	TRUE
ECU	0.17	135837	23751	8.04	TRUE
ENT	10.37	156873	1637252	8.38	FALSE
ENTl	5.38	130961	706518	8.68	FALSE

ENTl1	1.07	60192	64052	7.46	TRUE
ENTl2	1.30	137686	178335	8.95	TRUE
ENTl3	1.00	133503	134443	9.20	TRUE
ENTl5	1.14	143207	162553	8.60	TRUE
ENTl6a	0.87	186260	161651	8.04	TRUE
ENTm	4.99	184877	919750	8.16	FALSE
ENTm1	1.76	89895	158373	8.09	TRUE
ENTm2	1.11	220220	245358	8.39	TRUE
ENTm3	0.83	227632	186984	8.26	TRUE
ENTm5	0.77	263601	202095	8.00	TRUE
ENTm6	0.53	238519	124539	7.72	TRUE
EP	2.22	149740	331357	7.83	FALSE
EPI	0.56	272820	153417	7.88	FALSE
EPd	1.43	171372	243124	7.83	TRUE
EPv	0.80	108980	86442	7.80	TRUE
EW	0.02	228337	4510	7.52	TRUE
Eth	0.18	149106	27219	7.62	TRUE
FC	0.05	153312	7757	7.81	TRUE
FF	0.19	150037	28474	7.54	TRUE
FL	1.09	262898	292940	7.98	TRUE
FN	0.41	65806	27909	7.73	TRUE
FOTU	0.98	274695	273109	7.84	TRUE
FRP	0.78	156064	118746	8.25	FALSE
FRP1	0.21	68271	14147	7.76	TRUE
FRP2/3	0.17	152710	25250	8.67	TRUE
FRP5	0.30	191366	56384	8.27	TRUE
FRP6a	0.10	231237	22620	7.71	TRUE
FS	0.35	244193	85528	7.61	TRUE
GENd	1.14	179697	201385	7.85	FALSE
GENv	0.43	157238	67487	7.43	FALSE
GPe	1.50	122681	182855	7.47	TRUE
GPi	0.40	39636	15450	7.29	TRUE
GR	0.08	115708	8843	7.69	TRUE
GRN	2.37	46424	111805	7.62	TRUE
GU	1.44	206805	296482	8.06	FALSE
GU1	0.18	50137	8898	7.42	TRUE
GU2/3	0.32	199519	61651	8.39	TRUE
GU4	0.12	280415	33456	7.82	TRUE
GU5	0.45	219917	98790	8.04	TRUE
GU6a	0.34	252521	86919	7.88	TRUE
GU6b	0.03	207304	5977	7.70	TRUE

HATA	0.42	213232	90324	8.07	TRUE
HB	39.42	98567	3957059	7.63	FALSE
HEM	26.84	261836	7114499	7.92	FALSE
HIP	19.64	196988	3879152	8.15	FALSE
HPF	36.97	186226	6913768	8.20	FALSE
HY	12.30	192740	2372606	7.42	FALSE
IA	0.14	337542	48209	7.20	TRUE
IAD	0.09	199472	16866	7.97	TRUE
IAM	0.03	202317	6364	8.17	TRUE
IB	28.97	190464	5478473	7.71	FALSE
IC	3.62	258920	942315	7.65	FALSE
ICc	0.87	292873	256689	7.52	TRUE
ICd	1.06	286422	305750	7.73	TRUE
ICe	1.69	222509	377628	7.64	TRUE
IF	0.07	163954	11183	7.37	TRUE
IG	0.10	104289	9852	7.65	TRUE
ILA	0.68	222156	150569	8.33	FALSE
ILA1	0.14	106546	14727	8.41	TRUE
ILA2/3	0.12	228410	26630	8.70	TRUE
ILA5	0.25	241512	62111	8.40	TRUE
ILA6a	0.16	280776	45747	7.76	TRUE
ILA6b	0.01	225814	1941	7.50	TRUE
ILM	1.30	208015	269439	7.80	FALSE
IMD	0.13	198519	25838	8.12	TRUE
INC	0.06	146950	9562	7.62	TRUE
IO	0.43	115583	50936	7.66	TRUE
IP	0.71	143040	102646	7.73	TRUE
IPA	0.02	315992	5744	7.71	TRUE
IPC	0.05	182759	9807	7.79	TRUE
IPDL	0.03	229141	7289	7.71	TRUE
IPDM	0.01	367749	5377	7.66	TRUE
IPI	0.03	167522	4778	7.77	TRUE
IPL	0.05	116416	6159	7.50	TRUE
IPN	0.29	202946	58093	7.81	FALSE
IPR	0.06	229005	14585	8.07	TRUE
IPRL	0.01	153202	2244	7.61	TRUE
IRN	2.27	100349	227225	7.66	TRUE
IV	0.00	144510	565	7.92	TRUE
Isocortex	104.55	218930	22768158	8.07	FALSE
KF	0.16	153857	24349	7.56	TRUE
LA	0.65	207667	135222	7.99	TRUE

LAT	2.47	171070	421577	7.83	FALSE
LAV	0.24	40496	9844	7.67	TRUE
LC	0.01	241626	2311	7.64	TRUE
LD	0.80	177689	142213	8.17	TRUE
LDT	0.15	273766	42498	7.61	TRUE
LGd	0.57	187952	105618	7.83	FALSE
LGd-co	0.34	161994	55255	7.79	TRUE
LGd-ip	0.06	218178	12831	7.90	TRUE
LGd-sh	0.17	228224	37833	7.85	TRUE
LGv	0.34	157872	53313	7.45	TRUE
LH	0.29	227847	65063	7.73	TRUE
LHA	1.78	119495	212475	7.39	TRUE
LIN	0.05	121514	6161	7.86	TRUE
LING	0.11	234945	27370	7.81	TRUE
LM	0.07	143695	9374	7.43	TRUE
LP	1.00	176884	175339	7.83	TRUE
LPO	0.49	172191	84597	7.36	TRUE
LRN	0.45	55435	25237	7.90	FALSE
LRNm	0.42	59289	24911	7.91	TRUE
LS	2.43	214084	513076	7.87	FALSE
LSX	2.88	200444	573258	7.84	FALSE
LSc	0.45	137465	61700	9.05	TRUE
LSr	1.49	205847	306414	7.75	TRUE
LSv	0.49	297449	144312	7.36	TRUE
LZ	4.73	156540	757545	7.40	FALSE
MA	0.31	189241	58364	7.69	TRUE
MA3	0.01	217486	2545	7.32	TRUE
MARN	0.47	55310	25504	7.80	TRUE
MB	30.78	181205	5589113	7.64	FALSE
MBO	0.82	203470	166365	7.34	FALSE
MBmot	17.71	174863	3100951	7.61	FALSE
MBsen	5.78	257817	1516937	7.69	FALSE
MBsta	1.50	130714	199515	7.65	FALSE
MD	1.09	189932	203908	8.21	TRUE
MDRN	1.73	96360	162021	7.68	FALSE
MDRNd	0.91	105502	94925	7.68	TRUE
MDRNv	0.82	84957	68787	7.67	TRUE
ME	0.06	80262	5047	8.25	TRUE
MEA	1.71	205312	351304	7.62	TRUE
MED	1.59	194760	305901	8.11	FALSE
MEPO	0.03	311942	8550	7.36	TRUE

MEV	0.01	165800	1080	7.56	TRUE
MEZ	3.06	230825	708814	7.44	FALSE
MG	0.58	163972	96369	7.88	FALSE
MGd	0.14	202379	28665	8.04	TRUE
MGm	0.21	133902	28393	7.63	TRUE
MGv	0.22	176949	39058	7.86	TRUE
MH	0.27	317362	86595	7.97	TRUE
MM	0.45	229313	100737	7.28	FALSE
MMd	0.06	292632	17698	7.23	TRUE
MMl	0.18	186563	33914	7.32	TRUE
MMm	0.11	265859	29330	7.30	TRUE
MMme	0.06	241966	14260	7.23	TRUE
MMp	0.03	194865	5711	7.11	TRUE
MO	20.56	185666	3778046	8.34	FALSE
MOB	13.11	335148	4310375	7.70	TRUE
MOp	9.49	192495	1813402	8.27	FALSE
MOp1	1.30	72179	92595	7.71	TRUE
MOp2/3	2.88	215288	617830	8.53	TRUE
MOp5	2.64	201017	535332	8.30	TRUE
MOp6a	2.46	222547	539983	7.79	TRUE
MOp6b	0.17	142812	23228	7.38	TRUE
MOs	11.05	180149	1954995	8.42	FALSE
MOs1	2.24	67324	149180	8.00	TRUE
MOs2/3	2.93	198618	585414	8.78	TRUE
MOs5	3.67	201122	738237	8.37	TRUE
MOs6a	2.13	229055	481425	7.73	TRUE
MOs6b	0.07	112288	7528	7.18	TRUE
MPN	0.31	255818	78773	7.52	TRUE
MPO	0.43	209402	91641	7.42	TRUE
MPT	0.04	257488	10009	7.74	TRUE
MRN	4.45	90447	405470	7.48	TRUE
MS	0.34	210274	73857	7.48	TRUE
MSC	0.96	196178	189440	7.44	FALSE
MT	0.04	147912	5189	7.48	TRUE
MTN	0.96	230329	220177	7.93	FALSE
MV	1.49	195856	295614	7.80	TRUE
MY	26.16	92309	2508001	7.67	FALSE
MY-mot	14.48	95709	1416378	7.74	FALSE
MY-sat	0.20	67823	13256	7.72	FALSE
MY-sen	6.71	134307	892665	7.58	FALSE
ND	0.07	246059	17034	7.55	TRUE

NDB	0.61	185983	114226	7.39	TRUE
NI	0.07	223602	16423	7.43	TRUE
NLL	0.58	160676	93673	7.52	TRUE
NLOT	0.26	177053	45098	8.13	FALSE
NLOT1	0.09	95890	8438	7.44	TRUE
NLOT2	0.11	225904	25755	8.29	TRUE
NLOT3	0.05	207417	10555	7.90	TRUE
NOD	1.24	320298	402057	7.80	TRUE
NOT	0.18	200058	35724	7.65	TRUE
NPC	0.24	197208	47466	7.56	TRUE
NR	0.02	177644	4332	7.58	TRUE
NTS	0.70	206861	143402	7.49	TRUE
OLF	36.50	229090	8280728	7.92	FALSE
OP	0.05	277387	13345	7.64	TRUE
ORB	4.97	195001	953744	8.22	FALSE
ORBl	2.32	197370	448472	8.18	FALSE
ORBl1	0.31	91022	28305	7.65	TRUE
ORBl2/3	0.51	174260	88953	8.66	TRUE
ORBl5	1.01	222714	221687	8.16	TRUE
ORBl6a	0.44	232975	100672	7.75	TRUE
ORBl6b	0.04	200575	7917	7.74	TRUE
ORBm	1.15	176873	200956	8.21	FALSE
ORBm1	0.35	83674	29051	8.21	TRUE
ORBm2/3	0.25	180496	45393	8.61	TRUE
ORBm5	0.37	224868	85083	8.16	TRUE
ORBm6a	0.16	260232	40584	7.65	TRUE
ORBvl	1.48	203615	298936	8.28	FALSE
ORBvl1	0.31	127690	38886	7.56	TRUE
ORBvl2/3	0.43	183100	78740	8.63	TRUE
ORBvl5	0.52	234856	121413	8.34	TRUE
ORBvl6a	0.21	269112	55990	7.71	TRUE
ORBvl6b	0.01	252042	2010	7.51	TRUE
OT	3.16	215646	682016	7.83	TRUE
P	13.23	106228	1420952	7.56	FALSE
P-mot	4.57	118559	555498	7.64	FALSE
P-sat	2.69	99887	271949	7.51	FALSE
P-sen	3.03	135763	415793	7.51	FALSE
P5	0.25	104744	26410	7.44	TRUE
PA	0.85	220184	187690	8.36	TRUE
PAA	0.95	128062	121188	8.49	TRUE
PAG	3.86	199089	765725	7.50	FALSE

PAL	8.26	170138	1400284	7.51	FALSE
PALc	1.14	230324	261590	7.44	FALSE
PALd	1.89	106387	198010	7.46	FALSE
PALm	1.21	200126	244540	7.41	FALSE
PALv	3.01	198434	592839	7.59	FALSE
PAR	1.04	188507	194887	8.17	TRUE
PARN	1.77	76765	143111	7.62	TRUE
PAS	0.02	208690	4266	7.53	TRUE
PB	0.93	175697	163006	7.52	FALSE
PBG	0.04	145963	5210	7.57	TRUE
PC5	0.06	92800	5352	7.55	TRUE
PCG	0.47	222895	105536	7.43	TRUE
PCN	0.19	202231	37655	7.90	TRUE
PD	0.01	281994	1684	7.75	TRUE
PDTg	0.03	290730	8155	7.45	TRUE
PERI	0.66	116384	75833	8.89	FALSE
PERI1	0.21	55267	11831	8.00	TRUE
PERI2/3	0.26	136061	34867	9.20	TRUE
PERI5	0.13	141799	18124	8.63	TRUE
PERI6a	0.04	198632	8371	8.09	TRUE
PERI6b	0.01	165657	1753	7.93	TRUE
PF	0.39	201478	78144	7.64	TRUE
PFL	4.79	258270	1242245	8.00	TRUE
PG	0.75	211731	163181	7.75	TRUE
PGRN	0.79	58156	46438	7.59	FALSE
PGRNd	0.20	63314	12638	7.63	TRUE
PGRNl	0.59	55143	31776	7.56	TRUE
PH	0.58	232775	136493	7.39	TRUE
PHY	0.22	216432	46743	7.67	FALSE
PIL	0.14	169242	23931	7.41	TRUE
PIR	9.37	168134	1577945	8.15	TRUE
PL	1.97	191024	374070	8.40	FALSE
PL1	0.44	81040	35606	8.23	TRUE
PL2/3	0.34	200931	68506	8.78	TRUE
PL5	0.78	208161	162626	8.54	TRUE
PL6a	0.39	268406	104924	7.67	TRUE
PL6b	0.02	200597	3518	7.43	TRUE
PMd	0.11	263784	27917	7.61	TRUE
PMv	0.15	219456	31373	7.68	TRUE
PN	0.02	163631	2834	7.37	TRUE
PO	0.97	171962	166729	7.89	TRUE

POL	0.17	124143	20838	7.40	TRUE
POR	0.26	52854	13315	7.48	TRUE
POST	0.95	244868	232465	7.65	TRUE
PP	0.05	134811	6887	7.35	TRUE
PPN	0.76	80701	59378	7.40	TRUE
PPT	0.12	258777	30688	7.71	TRUE
PR	0.12	241427	28751	7.50	TRUE
PRC	0.14	194938	26173	7.63	TRUE
PRE	0.82	227087	188212	7.93	TRUE
PRM	4.33	254552	1109280	7.93	TRUE
PRNc	1.82	46742	85499	7.54	TRUE
PRNr	1.84	69871	129614	7.49	TRUE
PRP	0.20	219927	41835	7.67	TRUE
PRT	1.68	217131	365950	7.65	FALSE
PS	0.07	268207	19570	7.28	TRUE
PSTN	0.14	213392	29179	7.49	TRUE
PSV	0.89	126058	111522	7.40	TRUE
PT	0.18	196135	34388	8.25	TRUE
PTLp	2.06	243589	501851	7.97	FALSE
PVH	0.16	243308	38575	7.27	TRUE
PVHd	0.10	231953	24176	7.40	TRUE
PVR	1.57	229009	360990	7.40	FALSE
PVT	0.36	234301	84625	8.07	TRUE
PVZ	0.64	165294	106580	7.30	FALSE
PVa	0.04	222017	8101	7.24	TRUE
PVi	0.18	115978	20047	7.14	TRUE
PVp	0.11	213941	23276	7.39	TRUE
PVpo	0.11	203444	22531	7.20	TRUE
PYR	1.19	287656	339114	7.78	TRUE
Pa5	0.08	36355	2877	7.51	TRUE
PeF	0.18	191676	33534	7.37	TRUE
PoT	0.23	76551	17999	7.43	TRUE
ProS	1.10	144177	157347	8.59	TRUE
RAmb	0.59	183548	106923	7.71	FALSE
RCH	0.11	106344	11899	6.91	TRUE
RE	0.35	235947	82668	7.69	TRUE
RH	0.08	256492	19730	7.87	TRUE
RHP	16.91	173564	2942260	8.27	FALSE
RN	0.71	102095	70884	7.92	TRUE
RPA	0.05	25235	1234	7.73	TRUE
RPF	0.05	201438	10023	7.61	TRUE

RR	0.10	106315	10435	7.58	TRUE
RSP	9.49	246898	2322867	7.75	FALSE
RSPagl	2.17	240906	516241	7.93	FALSE
RSPagl1	0.63	110596	68359	7.98	TRUE
RSPagl2/3	0.56	328188	181490	8.07	TRUE
RSPagl5	0.62	269915	165317	7.93	TRUE
RSPagl6a	0.34	293917	98672	7.51	TRUE
RSPagl6b	0.03	124819	3559	7.00	TRUE
RSPd	3.50	227990	793059	7.78	FALSE
RSPd1	0.98	114679	112654	7.92	TRUE
RSPd2/3	0.84	309418	260220	7.86	TRUE
RSPd5	1.01	252980	254709	7.80	TRUE
RSPd6a	0.63	262481	163125	7.40	TRUE
RSPv	3.80	265315	1009113	7.58	FALSE
RSPv1	1.00	161414	161878	7.48	TRUE
RSPv2/3	0.88	397972	347343	7.45	TRUE
RSPv5	1.36	261212	356897	7.80	TRUE
RSPv6a	0.53	263597	136369	7.34	TRUE
RSPv6b	0.04	141102	5179	6.92	TRUE
RT	1.36	152134	207051	7.39	TRUE
SAG	0.08	164329	12715	7.34	TRUE
SBPV	0.08	291718	23476	7.34	TRUE
SCH	0.05	380041	17531	7.47	TRUE
SCO	0.01	83459	871	7.28	TRUE
SCdg	0.96	205187	198704	7.59	TRUE
SCdw	0.29	178077	51307	7.57	TRUE
SCig	1.70	275442	467427	7.66	TRUE
SCiw	1.67	241413	403220	7.65	TRUE
SCm	4.62	242246	1123737	7.65	FALSE
SCop	0.55	277068	150397	7.69	TRUE
SCs	1.99	274476	541677	7.76	FALSE
SCsg	0.98	322945	312577	7.80	TRUE
SCzo	0.47	166346	78320	7.67	TRUE
SF	0.42	124535	51630	7.32	TRUE
SFO	0.01	239228	3466	7.62	TRUE
SG	0.01	277829	2890	7.54	TRUE
SGN	0.15	177387	26906	7.72	TRUE
SH	0.03	260164	7354	8.16	TRUE
SI	2.70	196544	529964	7.58	TRUE
SIM	4.98	245485	1243957	7.88	TRUE
SLC	0.02	148638	2981	7.70	TRUE

SLD	0.03	200749	6491	7.55	TRUE
SMT	0.26	179451	46446	7.94	TRUE
SNc	0.16	165043	26950	7.76	TRUE
SNr	1.26	150506	191887	7.61	TRUE
SO	0.04	106489	3977	7.18	TRUE
SOC	0.64	76128	49104	7.45	FALSE
SOCl	0.25	97443	24560	7.41	TRUE
SPA	0.10	173222	15855	7.63	TRUE
SPF	0.17	143438	23594	7.48	FALSE
SPFm	0.06	207766	11566	7.55	TRUE
SPFp	0.11	107859	11580	7.33	TRUE
SPIV	0.62	89845	54371	7.66	TRUE
SPVC	1.51	153116	228757	7.54	TRUE
SPVI	1.52	112277	168025	7.51	TRUE
SPVO	0.85	79439	68967	7.53	TRUE
SS	28.28	241780	6814304	7.88	FALSE
SSp	20.77	248060	5136722	7.86	FALSE
SSp-bfd	5.38	261559	1395829	7.80	FALSE
SSp-bfd1	0.83	100700	84175	8.16	TRUE
SSp-bfd2/3	1.23	257147	315910	8.20	TRUE
SSp-bfd4	1.11	388754	429729	7.37	TRUE
SSp-bfd5	0.96	235000	223555	7.80	TRUE
SSp-bfd6a	1.09	290314	318157	7.46	TRUE
SSp-bfd6b	0.13	156285	19758	7.28	TRUE
SSp-ll	1.99	241891	477202	7.97	FALSE
SSp-ll1	0.30	84747	25568	7.81	TRUE
SSp-ll2/3	0.51	255889	130698	8.27	TRUE
SSp-ll4	0.22	383253	86068	7.45	TRUE
SSp-ll5	0.46	230792	104776	8.18	TRUE
SSp-ll6a	0.45	273132	123423	7.66	TRUE
SSp-ll6b	0.04	146775	6632	7.15	TRUE
SSp-m	5.33	233815	1243897	7.88	FALSE
SSp-m1	0.77	80083	60906	8.18	TRUE
SSp-m2/3	1.23	234457	287487	8.36	TRUE
SSp-m4	0.81	361746	292918	7.46	TRUE
SSp-m5	1.04	229970	239606	7.80	TRUE
SSp-m6a	1.38	251670	345248	7.48	TRUE
SSp-m6b	0.09	170703	15074	7.33	TRUE
SSp-n	2.61	250954	652369	7.82	FALSE
SSp-n1	0.40	96781	39037	8.30	TRUE
SSp-n2/3	0.54	246505	134398	8.29	TRUE

SSp-n4	0.49	383127	186522	7.41	TRUE
SSp-n5	0.46	237252	109301	7.73	TRUE
SSp-n6a	0.67	265791	179973	7.41	TRUE
SSp-n6b	0.05	120110	5207	7.26	TRUE
SSp-tr	1.16	253608	295608	7.96	FALSE
SSp-tr1	0.21	94106	19556	7.76	TRUE
SSp-tr2/3	0.34	279337	94025	8.17	TRUE
SSp-tr4	0.12	396041	48360	7.45	TRUE
SSp-tr5	0.28	250478	70610	8.12	TRUE
SSp-tr6a	0.18	296481	55035	7.66	TRUE
SSp-tr6b	0.03	198173	5642	7.17	TRUE
SSp-ul	3.23	243430	779130	7.88	FALSE
SSp-ul1	0.45	88103	39774	7.95	TRUE
SSp-ul2/3	0.78	251918	196882	8.28	TRUE
SSp-ul4	0.45	391033	174267	7.36	TRUE
SSp-ul5	0.65	230637	150938	7.92	TRUE
SSp-ul6a	0.82	257147	205788	7.51	TRUE
SSp-ul6b	0.07	120094	7851	7.22	TRUE
SSp-un	1.07	246022	262257	7.84	FALSE
SSp-un2/3	0.26	251207	66126	8.19	TRUE
SSp-un4	0.16	369649	57604	7.39	TRUE
SSp-un5	0.21	233549	49371	7.84	TRUE
SSp-un6a	0.26	273096	69406	7.47	TRUE
SSp-un6b	0.02	125526	2689	7.21	TRUE
SSs	7.47	222227	1662944	7.95	FALSE
SSs1	1.19	78772	93982	8.19	TRUE
SSs2/3	1.68	222464	371284	8.38	TRUE
SSs4	0.93	313654	292420	7.63	TRUE
SSs5	1.78	231191	413173	7.85	TRUE
SSs6a	1.70	268504	457006	7.64	TRUE
SSs6b	0.17	193267	32828	7.57	TRUE
STN	0.16	204664	32868	7.43	TRUE
STR	37.75	221960	8360744	7.79	FALSE
STRd	22.19	214194	4719098	7.83	FALSE
STRv	7.07	262506	1862334	7.75	FALSE
SUB	1.87	195125	366965	8.26	TRUE
SUM	0.20	188415	38274	7.36	TRUE
SUT	0.20	146272	28790	7.56	TRUE
SUV	0.29	98238	28561	7.70	TRUE
Su3	0.03	221830	7187	7.58	TRUE
SubG	0.02	209493	4313	7.37	TRUE

TEa	2.56	208633	529080	8.17	FALSE
TEa1	0.54	73303	38816	8.05	TRUE
TEa2/3	0.54	222698	119087	8.47	TRUE
TEa4	0.24	264608	63104	7.97	TRUE
TEa5	0.74	234382	173963	8.14	TRUE
TEa6a	0.42	281401	117694	7.94	TRUE
TEa6b	0.08	195662	14626	7.62	TRUE
TH	16.66	182833	3050178	7.88	FALSE
TM	0.10	156395	15849	7.45	FALSE
TMd	0.02	248213	5798	7.55	TRUE
TMv	0.08	127454	9932	7.35	TRUE
TR	1.17	109374	129315	8.82	TRUE
TRN	0.59	87415	53351	7.70	TRUE
TRS	0.25	213065	51286	7.27	TRUE
TT	1.16	134682	154945	8.45	FALSE
TTd	0.60	152218	91132	8.63	TRUE
TTv	0.56	114415	62493	8.00	TRUE
TU	0.42	131304	55874	7.22	TRUE
UVU	2.14	283183	607901	7.74	TRUE
V	0.29	111356	32131	7.87	TRUE
VAL	0.67	165137	110717	8.01	TRUE
VCO	0.85	114228	96826	7.47	TRUE
VENT	4.05	165372	670591	7.79	FALSE
VERM	18.19	277678	5144896	7.82	FALSE
VII	0.70	87558	62537	7.90	TRUE
VIS	11.90	254182	3014946	7.90	FALSE
VISC	1.93	191332	372807	8.12	FALSE
VISC1	0.32	56669	17836	7.81	TRUE
VISC2/3	0.44	194294	84816	8.47	TRUE
VISC4	0.14	264795	36462	7.81	TRUE
VISC5	0.57	203056	115661	8.13	TRUE
VISC6a	0.44	244749	108020	7.87	TRUE
VISC6b	0.05	216388	9744	7.58	TRUE
VISa	1.21	245126	296350	8.03	FALSE
VISa1	0.24	99844	24371	7.91	TRUE
VISa2/3	0.34	283293	96835	8.22	TRUE
VISa4	0.13	349883	43964	7.69	TRUE
VISa5	0.29	246667	71937	8.10	TRUE
VISa6a	0.18	301426	53629	7.53	TRUE
VISa6b	0.03	161786	5232	7.06	TRUE
VISal	0.63	244091	152198	7.82	FALSE

VISal1	0.12	107242	12426	7.91	TRUE
VISal2/3	0.15	273558	42246	8.06	TRUE
VISal4	0.08	338778	27883	7.53	TRUE
VISal5	0.16	247123	40153	7.82	TRUE
VISal6a	0.09	294199	26990	7.47	TRUE
VISal6b	0.02	163035	2973	7.13	TRUE
VISam	0.67	244459	163920	8.05	FALSE
VISam2/3	0.17	290464	48896	8.32	TRUE
VISam5	0.19	260310	49412	8.06	TRUE
VISam6a	0.11	306590	32407	7.38	TRUE
VISl	1.07	245831	260270	7.82	FALSE
VISl1	0.21	113143	23659	8.14	TRUE
VISl2/3	0.24	280181	64913	8.09	TRUE
VISl4	0.15	335498	49797	7.53	TRUE
VISl5	0.26	253934	65176	7.77	TRUE
VISl6a	0.18	297766	52017	7.51	TRUE
VISl6b	0.03	120086	4008	7.14	TRUE
VISli	0.41	244731	99939	7.85	FALSE
VISli1	0.08	103780	8546	7.93	TRUE
VISli2/3	0.09	280937	25619	8.12	TRUE
VISli4	0.04	343744	14683	7.58	TRUE
VISli5	0.12	264836	31430	7.84	TRUE
VISli6a	0.06	294890	18483	7.59	TRUE
VISli6b	0.01	135240	1686	7.13	TRUE
VISp	6.27	268271	1685308	7.84	FALSE
VISp1	1.53	131139	199851	8.13	TRUE
VISp2/3	1.63	313165	510362	8.06	TRUE
VISp4	0.83	381427	317664	7.48	TRUE
VISp5	1.29	272634	352025	7.74	TRUE
VISp6a	0.87	326089	283895	7.50	TRUE
VISp6b	0.15	139970	20417	7.15	TRUE
VISpl	0.79	231319	183159	8.00	FALSE
VISpl1	0.27	127614	34411	8.12	TRUE
VISpl2/3	0.18	296102	53378	8.11	TRUE
VISpl4	0.03	299103	8687	7.86	TRUE
VISpl5	0.21	264417	55429	7.97	TRUE
VISpl6a	0.09	312511	28991	7.66	TRUE
VISpl6b	0.01	167221	1023	7.39	TRUE
VISpm	0.91	251740	226950	8.06	FALSE
VISpm1	0.21	106932	22490	8.05	TRUE
VISpm2/3	0.24	291973	70606	8.35	TRUE

VISpm4	0.08	357940	29770	7.75	TRUE
VISpm5	0.23	266905	62104	8.00	TRUE
VISpm6a	0.12	325516	38397	7.44	TRUE
VISpm6b	0.02	165472	3738	7.09	TRUE
VISpor	1.14	214409	243345	8.15	FALSE
VISpor1	0.31	89078	27317	8.19	TRUE
VISpor2/3	0.30	255561	75160	8.37	TRUE
VISpor4	0.05	325593	15324	7.78	TRUE
VISpor5	0.31	244282	76369	8.12	TRUE
VISpor6a	0.15	297259	44149	7.83	TRUE
VISpor6b	0.03	144501	3955	7.48	TRUE
VISrl	0.85	242904	204173	7.89	FALSE
VISrl1	0.17	114461	19415	8.09	TRUE
VISrl2/3	0.22	279158	60357	8.11	TRUE
VISrl4	0.12	336683	40518	7.54	TRUE
VISrl5	0.20	229808	45142	7.93	TRUE
VISrl6a	0.12	295175	35483	7.44	TRUE
VISrl6b	0.02	145000	3466	7.07	TRUE
VLPO	0.05	79246	4162	7.21	TRUE
VM	0.77	150946	113912	7.78	TRUE
VMH	0.43	235542	101503	7.46	TRUE
VMPO	0.03	106643	3481	6.91	TRUE
VNC	2.64	146047	384944	7.78	FALSE
VP	2.37	178903	420971	7.74	FALSE
VPL	0.78	138886	109531	7.58	TRUE
VPLpc	0.08	152514	12055	7.84	TRUE
VPM	1.32	207796	269500	7.74	TRUE
VPMpc	0.18	163308	29930	8.03	TRUE
VTA	0.37	125981	47089	7.72	TRUE
VeCB	0.08	41650	3262	7.58	TRUE
XII	0.23	154182	34772	8.18	TRUE
Xi	0.07	261104	18035	7.56	TRUE
ZI	1.46	187241	273893	7.43	FALSE
grey	376.99	204624	77265992	7.94	FALSE
sAMY	3.29	210943	700324	7.63	FALSE
y	0.02	175498	2881	7.63	TRUE

תקציר

מוח של עכבר נחקר בצורה יסודית ביותר אך ההערכות לגבי צפיפות התאים והשונות שלהם אינה בת השגה באזורים רבים במוח ולגבי אזורים נוספים קיימת אקסטרפולציה בלבד. בנוסף לכך, לא קיימת בספרות הערכה לגבי השונות בנפח ובצפיפות התאים בין מוחות יחידניים. מכון המחקר ע"ש דיוויד אלן ייצר סדרת תמונות חתכים עבור מאות מוחות שלמים אשר יכולים לשמש כניסיון ראשון לענות על שאלות אלה. כחלק מהמחקר הזה פיתחתי מערכת לזיהוי אובייקטים מבוססת רשת נוירונים עמוקה (DNN) אשר משתמשת בתמונות אוטופלורסצנטיות לזיהוי גרעיני התאים בתוך המוח, ואת אף באזורים הצפופים ביותר כדוגמת Dentate gyrus. הפעלתי את המערכת על 537 סדרות חתכים של מוחות מתוך Allen Brain Connectivity Project. הנתונים שהתקבלו מאפשרים לנתח נתונים מתוך מוחות של זכרים ונקבות מזן C57BL/6J ו-FVB.CD1 כדי לבצע לזהות הבדלים בין-מיניים וכן הבדלים בין הזנים.

האוניברסיטה הפתוחה
המחלקה למתמטיקה ומדעי המחשב

ספירה אוטומטית של תאי מוח עכבר שאינה דורשת צביעה מגלה תכונות נוירו-אנטומיות תלויות מין וזן

מאת

דוד אלקינד

התזה מוגשת כחלק מדרישות תואר מוסמך במדעים
במדעי המחשב
המחלקה למתמטיקה ולמדעי המחשב

התזה הוכנה בהנחייתם של
פרופ' נעם שנטל (האוניברסיטה הפתוחה)
פרופ׳ עמית צייזל (הטכניון)

אדר ב׳ תשפ״ב