Supplemental Document

Biomedical Optics EXPRESS

Large-scale, cell-resolution volumetric mapping allows layer-specific investigation of human brain cytoarchitecture: supplement

IRENE COSTANTINI,[1,2,3,9,*] **GIACOMO MAZZAMUTO,**[1,3,9] **MATTEO ROFFILLI,**[4] **ANNUNZIATINA LAURINO,**[1,8] **FILIPPO MARIA CASTELLI,**[1,5] **MATTIA NERI,**[4] **GIOVANNI LUGHI,**[4] **ANDREA SIMONETTO,**[4] **ERICA LAZZERI,**[1] **LUCA PESCE,**[1,5] **CHRISTOPHE DESTRIEUX,**[6] **LUDOVICO SILVESTRI,**[1,3,5] **VALERIO CONTI,**[7] **RENZO GUERRINI,**[7] **AND FRANCESCO SAVERIO PAVONE**[1,3,5]

[1] *European Laboratory for Non-Linear Spectroscopy (LENS), University of Florence, Sesto Fiorentino, Italy*
[2] *Department of Biology, University of Florence, Italy*
[3] *National Institute of Optics (INO), National Research Council (CNR), Sesto Fiorentino, Italy*
[4] *Bioretics srl, Cesena, Italy*
[5] *Department of Physics, University of Florence, Italy*
[6] *UMR 1253, iBrain, Université de Tours, Inserm, Tours, France*
[7] *Pediatric Neurology, Neurogenetics and Neurobiology Unit and Laboratories, A. Meyer Children's Hospital, University of Florence, Florence, Italy*
[8] *Present address: Department of Neurofarba, Section of Pharmacology and Toxicology, University of Florence, Italy*
[9] *These authors contributed equally to this work*
[*] *costantini@lens.unifi.it*

Supplement DOI: https://doi.org/10.6084/m9.figshare.14338433

Parent Article DOI: https://doi.org/10.1364/BOE.415555

Large-scale, cell-resolution volumetric analysis allows layer-specific investigation of human brain cytoarchitecture: supplemental document

SWITCH/TDE sample clearing. To demonstrate the widespread applicability of the SWITCH/TDE protocol different samples were cleared. Supplementary figure S1 shows 4 autoptic pieces from healthy subjects and 5 blocks from diseased patients of various ages, kept in formalin for different amounts of time before and after the clearing procedure.

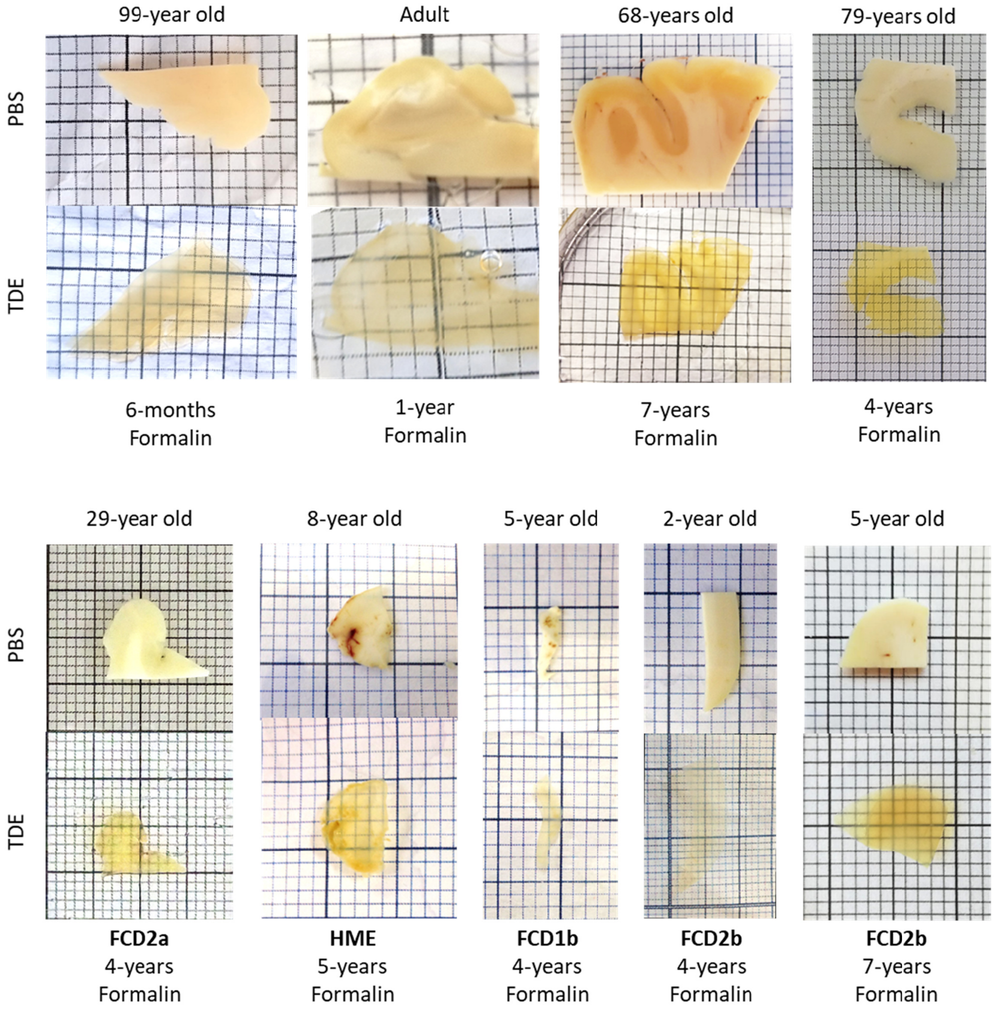


Fig. S1. Picture of various samples before (PBS) and after (TDE) the clearing protocol.
Acronym: HME (Hemimegalencephaly), FCD (Focal cortical dysplasia).

Immunostaining protocol optimization. Trials with different temperatures (37°C, @RT, and 4°C) and time of incubation (1 or 2 days) were performed in order to identify the best condition for optimal immunostaining (Supplementary figure S2). For each condition, images were acquired with the TPFM maintaining the same PMT gain and laser power. Signal to

Background (S/B) analysis was performed with Fiji (http://fiji.sc/Fiji) to assess the best contrast: mean intensity of 10 square of 25px were analyzed for each condition, media and standard deviation were then calculated using OriginPro 9.0 (OriginLab Corporation). The protocol with the highest signal amplification corresponds to 4°C of incubation temperature and 24 hours of incubation time of the primary antibody.

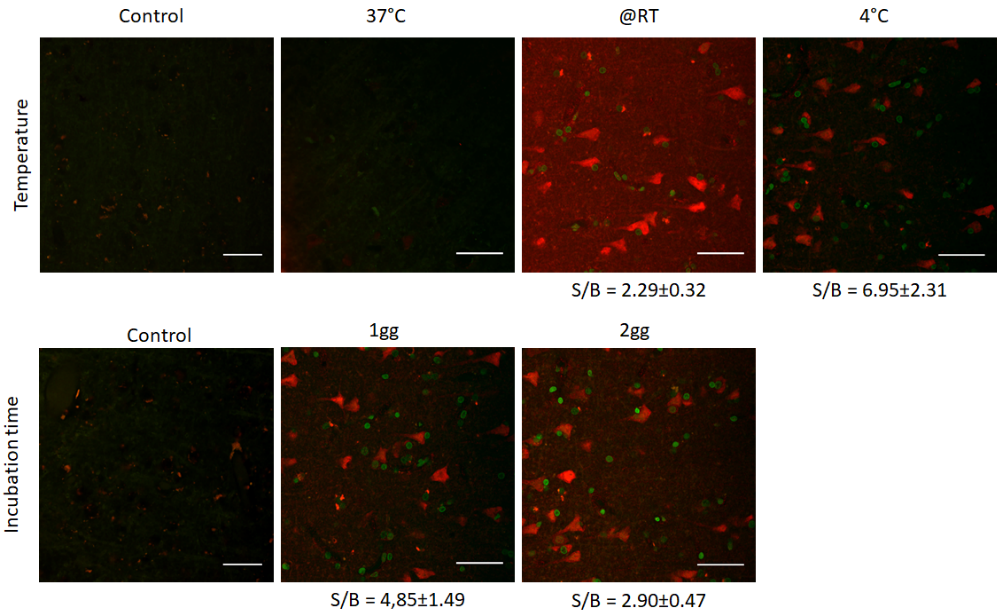


Fig. S2. Representative two-photon images of tissue stained with anti-NeuN antibody (in red) and DAPI (in green) at different temperatures and incubation times. Scale bar = 50 µm.

Image contrast variability. Biological specimens present heterogeneity in the staining process caused by different variables. There are intrinsic factors, as the composition of the tissue (e.g more or less myelinization) and external processes, like fixation and storage conditions, that change the outcome of the labeling. In addition to that, mesoscopic reconstructions are subjected to alteration of light absorption due to the presence of various tissue components that can introduce higher light absorption or scattering (eg. blood vessels). All these aspects lead to an image contrast variability between different specimens and inside the same samples through the scanned surface and the depth as shown in supplementary figure S3.

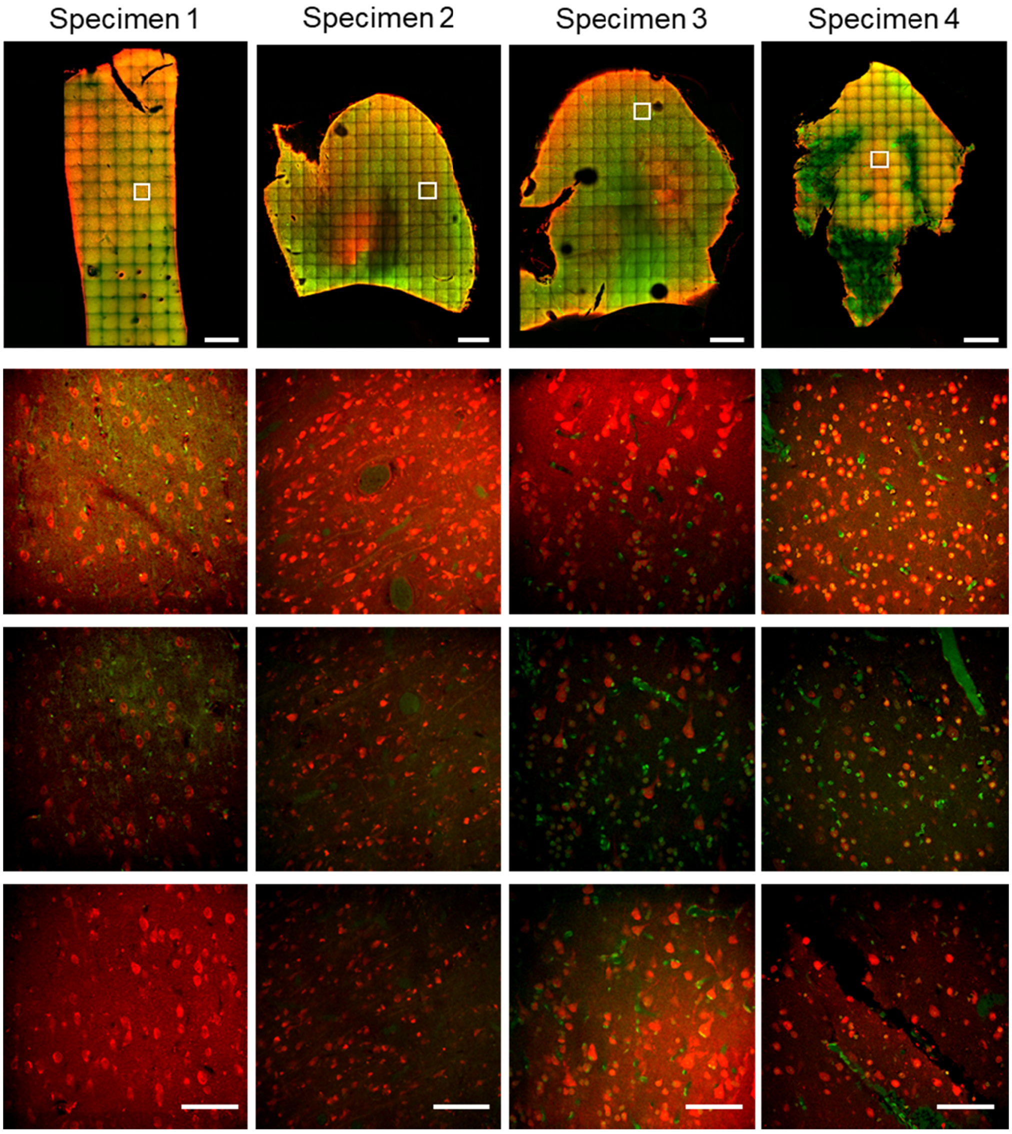


Fig. S3. First row: a representative middle plane (≈ 200 µm) of the mesoscopic reconstruction of the four analyzed human brain specimens obtained with TPFM. Scale bar = 1mm. Bottom rows: magnified insets of each specimen showing the native resolution of the acquisition. Images were selected from the top, middle plane and bottom of the stacks indicated with a white box in the first row. Scale bar = 50 µm.

Traditional method cell segmentation assessment. The main consequence of the high image contrast variability in mesoscopic measurements is the difficulty in the automatic analysis of the data. Indeed, traditional approaches are not sufficient to successfully identify and segment the stained neurons. To demonstrate the necessity of using a machine learning-based approach to analyze our data, we evaluated images of a representative stack at different depths (i.e 40 µm, 200 µm and 400 µm) using various tools provided by Fiji. In supplementary figure S4 two threshold-based methods and a color clustering one are compared with the 2.5D approach. The first method estimates the cell-occupied area with a brightness threshold, the obtained binary mask is then refined by consecutive morphological erosion and dilation, followed by morphological hole filling to ensure topological regularity of the detected objects. An object dimension filter is then applied to the segmentation mask to reject spurious noise-induced detections.

In the second method, in addition to the threshold-based estimation of the area occupied by cells, we calculate the brightness maxima and, assuming each maximum corresponds to a cell, we partition the image space applying a watershed algorithm directly to the brightness values (not the Euclidean Distance Transform), the two obtained segmentation maps are then combined with a logical AND, and lastly morphological hole filling and dimension-based spurious shape rejection are applied.

The third method performs an estimation of the cell-occupied areas by K-means color clustering choosing a number of 2 clusters, the clustering map is then morphologically refined with erosion, dilation and hole filling, a second segmentation map is obtained via the same watershed algorithm of the previous method and the two maps are combined using an AND logical operation, lastly spurious segmentations are rejected using an object size filter.

The three methods struggle to generalize with respect to brightness and contrast variability, yielding to generally poor segmentation results in challenging imaging conditions like the ones that characterize the initial part of our dataset (Fig. S4, 40 µm). Subpar detection performance is observed even in the other parts of the dataset, with many medium and low-intensity objects being completely lost in segmentation (Fig. S4, 200 and 400 µm).

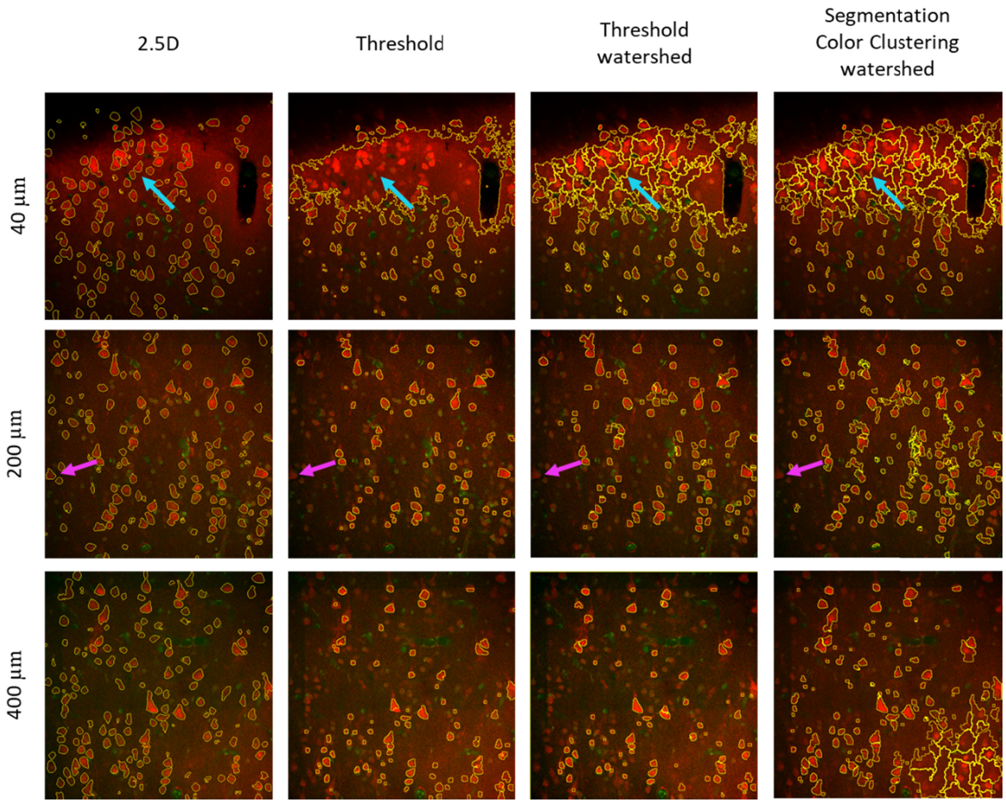
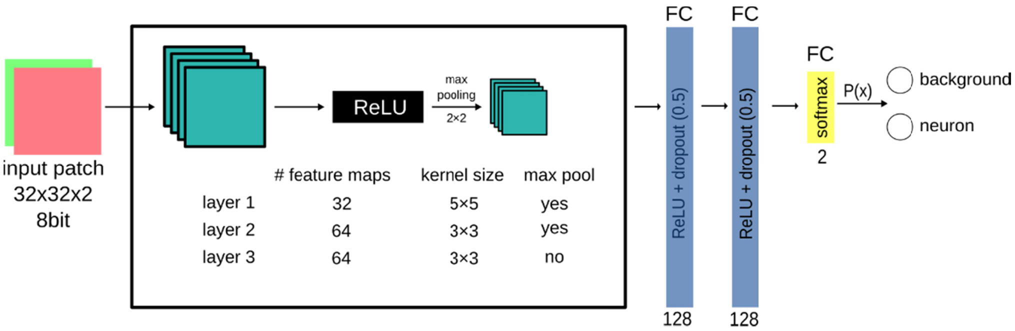


Fig. S4. Images selected from different planes (i.e 40 µm, 200 µm and 400 µm) of a representative stack analyzed with the 2.5D approach and different traditional approaches (obtained using Fiji) for segmentation assessment comparison. Cyan and magenta arrows indicate areas of wrong segmentations either for false positive or false negative recognition of traditional methods compared to the 2.5D approach.

Automatic cell counting and volumetric segmentation. We used a Convolutional Neural Network (CNN) for pixel-based classification followed by an analytical reconstruction of 3D polygonal meshes. The network uses information jointly from the red and green channels (i.e. neurons vs nuclei and tissue autofluorescence) to assign to each pixel a probability of belonging to either the neuron or the background class (Supplementary Figure S5). We adopted a pure 2-class fully convolutional CNN that transforms the multichannel source image into a new grayscale one, the so-called probability heatmap.



	# feature maps	kernel size	max pool
layer 1	32	5×5	yes
layer 2	64	3×3	yes
layer 3	64	3×3	no

Fig. S5. The CNN architecture. Block scheme of the architecture of the CNN with 3 convolutional layers and 3 Fully Connected layers.

In detail, each 2D images of the stack are processed independently by the neural network producing a probabilistic map of neuronal presence. Then, the heatmaps of all the acquired plans are reassembled back into a 3D stack. Instance semantic segmentation, based on an iso-surface finding algorithm, is then performed, with a statistical acceptance threshold of 0.5, to the heatmap volume in order to extract the three-dimensional surfaces of each uniquely identified polyhedron. The final results of this approach is a vectorial map describing the surfaces of each neuron in the 3D space (Supplementary Figure S6 and Supplementary video 1).

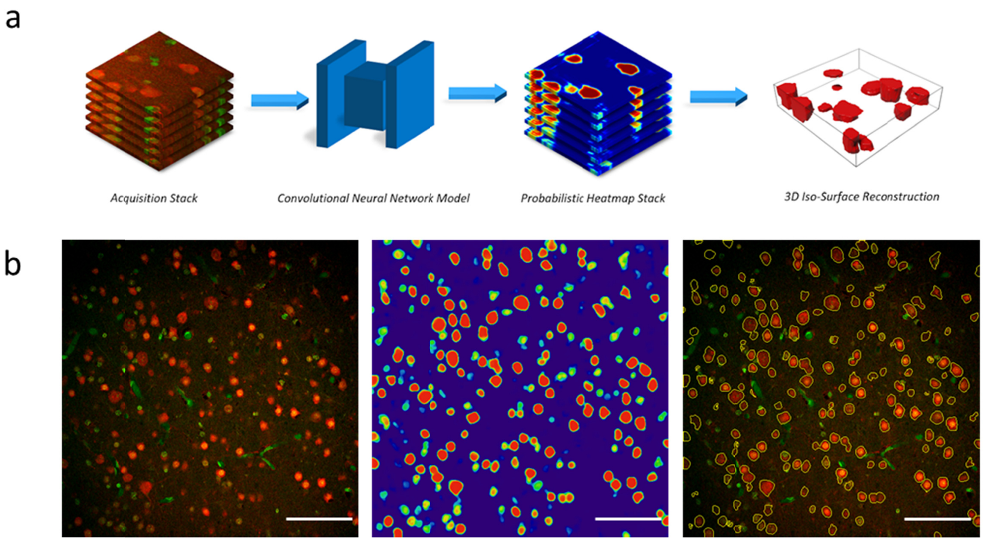


Fig. S6. (a) Neuronal segmentation workflow of the 2.5D approach. (b) A representative image undergoing the CNN analysis. From the native image to cells contour segmentation. Scale bar = 100 µm

The use of a relatively light model (in terms of number of free parameters) allowed us to obtain good segmentation results with advantages on both computational costs and annotation requirements: fast inference times allowed us to obtain results in almost-real-time (with respect to the acquisition time at the microscope) while the number of trainable parameters made it possible to train the model in a supervised fashion using a manageable amount of manually annotated data necessary for the ground truth. Moreover, the use of a 2D CNN segmentation network enables us to cover an annotation effort that would be infeasible if we were to train a 3D CNN model instead: this is because the necessary volumetric ground truth would geometrically increase the number of the required manually annotated voxels with respect to the planar labeling that was used for our 2D CNN model.

Convolutional Neural Networks (CNN) parameters. The 2.5D CNN was implemented using the Aliquis® software ecosystem (https://www.bioretics.com/aliquis) version 2.4.3 with TensorFlow backend (https://www.tensorflow.org). Aliquis® is available online free of charge in capped mode.

The network architecture is as follows:

input 32x32x2 8-bit (promoted to float4 and remapped in range [0 1])

conv (relu) 32 5x5 + max pooling 2x2

conv (relu) 64 3x3 + max pooling 2x2

conv 64 3x3 (relu)

fc (relu) 128 + dropout (0.5)

fc (relu) 128 + dropout (0.5)

fc 2 (softmax)

The network parameters are as follows:

Layer name (type)	**Output**	**N. params**
conv2d_1 (Conv2D)	32	1632
max_pooling2d_1 (MaxPool 2)	32	0
conv2d_2 (Conv2D)	64	18496
max_pooling2d_2 (MaxPool 2)	64	0
conv2d_3 (Conv2D)	64	36928
conv2d_4 (Conv2D)	128	131200
dropout_1 (Dropout)	128	0
conv2d_5 (Conv2D)	128	16512
dropout_2 (Dropout)	128	0
output (Conv2D)	2	256

Total params:	205,024
Trainable params:	205,024
Non-trainable params:	0

Optimizer hyper-parameters are as follows:

type: SGD (mini-batch stochastic)
batch size: 256
epochs: 300
scale: 0.003921568627 (= 1/255)
learning rate: 0.01
weight decay: 0.00001
momentum: 0.9
loss: infogain_categorical_crossentropy
infogain weight matrix: 0.7; 0; 0 ;1

The datasets extracted from the raw images are as follows:

Summary per-image dataset:

training = 112 (80%); validation = 14 (10%); test = 14 (10%)

Summary per-patch dataset (number of 32x32x2 samples):

Total = 2'293'760; background 0 = 2'147'474 (94%); neuron 1 = 146'286 (6%)

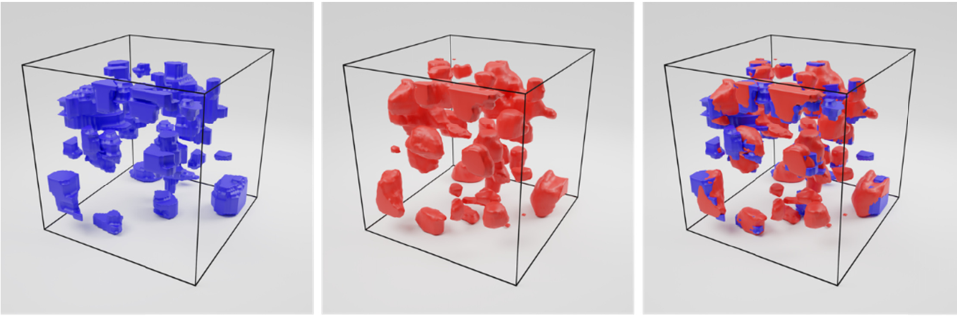
Detailed per-patch dataset (number of 32x32x2 samples):

Training set: background 0 = 1'717'979; neuron 1 = 117'029

Validation set: background 0 = 214'747; neuron 1 = 14'629

Only the training dataset has been data-augmented.

2.5D statistical assessment. The statistical assessment of the 2.5D performance was determined by analyzing four representative stacks of 100 x 100 x 100 μm^3, one for each specimen and a representative stack of 100 x 100 x 450 μm^3 to determine the accuracy along the full depth. Each stack was independently manually annotated by an operator and automatically segmented by the 2.5D approach. The manual segmentation has been performed on LAIRA® (https://laira.bioretics.com) and stored in the Ximage open-source format (https://github.com/bioretics/ximage). LAIRA® is available free of charge in trial mode. Supplementary Figure S7 shows the comparison between the manual annotations and the automatic reconstruction for one of these sub-volumes.



GT 2.5D Merge

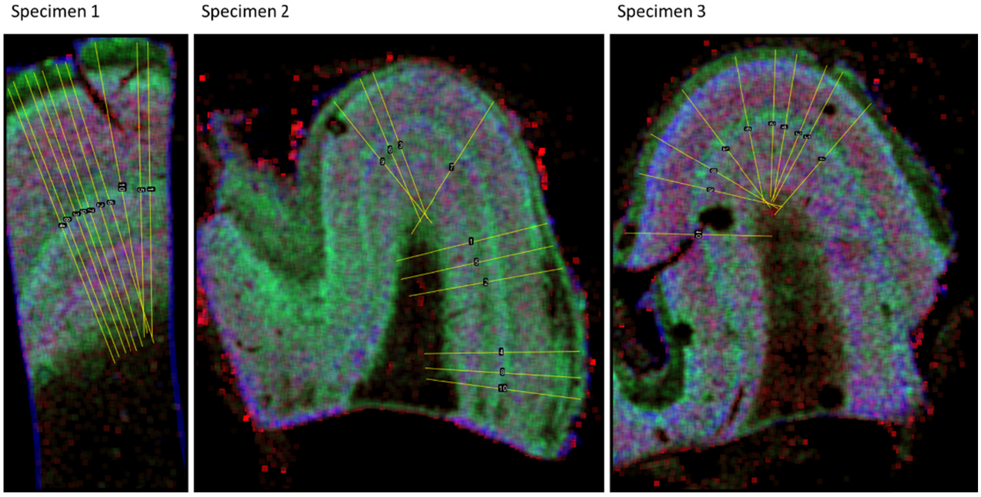
Fig. S7. 3D representation of the neurons of a stack manually annotated by the operator (in blue), automatically identified by the 2.5D approach (in red), and the superposition of the two.

We performed the evaluation for the two adopted approaches: cell counting and cell volumetric evaluation. The reported values have been computed on a macro statistics basis, i.e. firstly the average of all data of one single specimen is computed and then the average and standard deviation on the five specimens is derived (2.5D_assessment.zip Dataset1).

The total number of neurons found by the 2.5D is 442 over the 474 neurons annotated in the GT and the volumetric evaluation of the neurons is characterized by a precision of 91 ±7% and a recall rate of 96 ± 3%.

To verify the accuracy of the 2.5D along with depth we calculated the performance of the method dividing the 100 x 100 x 450 μm^3 stack in five fractions of 0-100; 101-200; 201-300; 301-400; 401-450 μm depth. The cell counting obtained by the 2.5D approach respect to the GT for each of these substacks are: 38/42, 57/65, 56/43, 61/64, 22/40 while the precision and the recall rate of the volumetric evaluation are respectively 91% and 97%. The results are in line with the overall assessment of the 2.5D but considering the limitation of analyzing just one representative stack, to better describe the accuracy of 2.5D in counting and volume evaluation, we performed a more detailed analysis on all the samples and along with the cortical depth as described in the next paragraph.

Depth accuracy quantification. To describe the variability of the 2.5D along with depth, we plotted the z-profiles across the cortex along 10 different lines that were drawn orthogonally to the cortical layers in specimens 1, 2, and 3 using Fiji. To avoid the distribution artifacts introduced by the imaging (e.g alteration of the surface and the bottom of the sample due to the fact that the tissue is not completely flat), for each sample we selected the middle planes (δ depth≈200µm) that present intact surfaces. Supplementary Figures S8 shows for each specimen the 10 selected lines. Then, we calculated and plotted the mean value and the one standard deviation of each profile along Z (Supplementary Figures S9, S10, S11). The stability of the profile along with the cortical depth and the stack depth proves that the 2.5D evaluation is stable between different samples, in various regions of the cortex, and along with the stack depth.



Specimen 1

Specimen 2

Specimen 3

Fig. S8. The figure shows the 10 selected lines used for the analysis for specimens 1, 2, and 3 drown on top of the maps of figure 5d.

Specimen 1

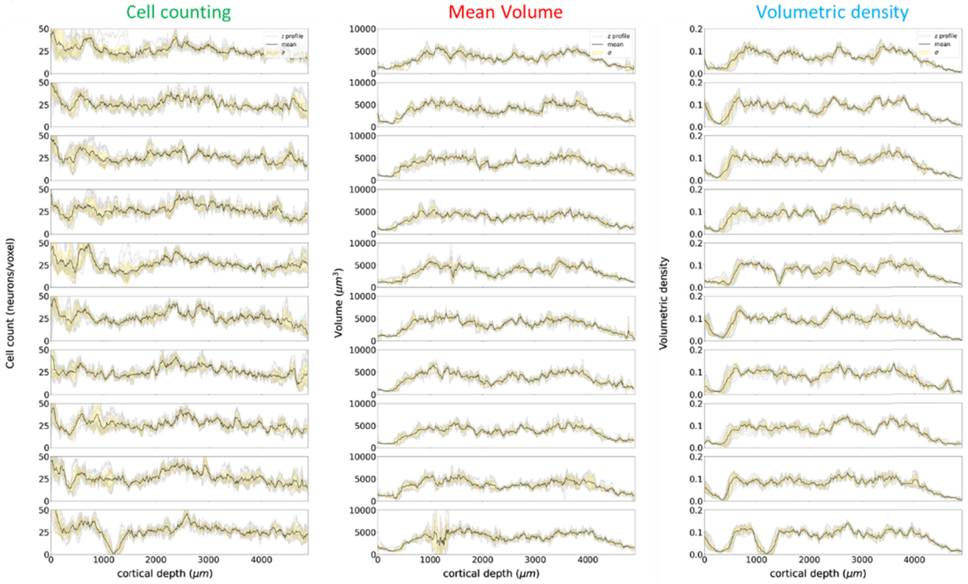


Fig. S9. Plots of cortical depth profiles (i.e. along Z) are shown together with the mean value and the one standard deviation for each of the 10 lines of specimen 1 showed in Fig S7.

Specimen 2

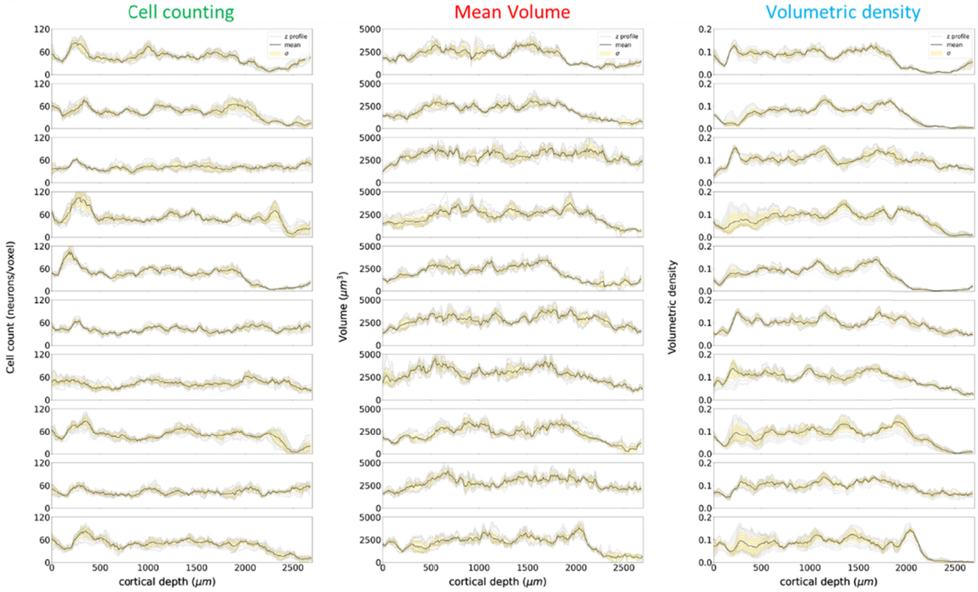


Fig. S10. Plots of cortical depth profiles (i.e. along Z) are shown together with the mean value and the one standard deviation for each of the 10 lines of specimen 2 showed in Fig S7

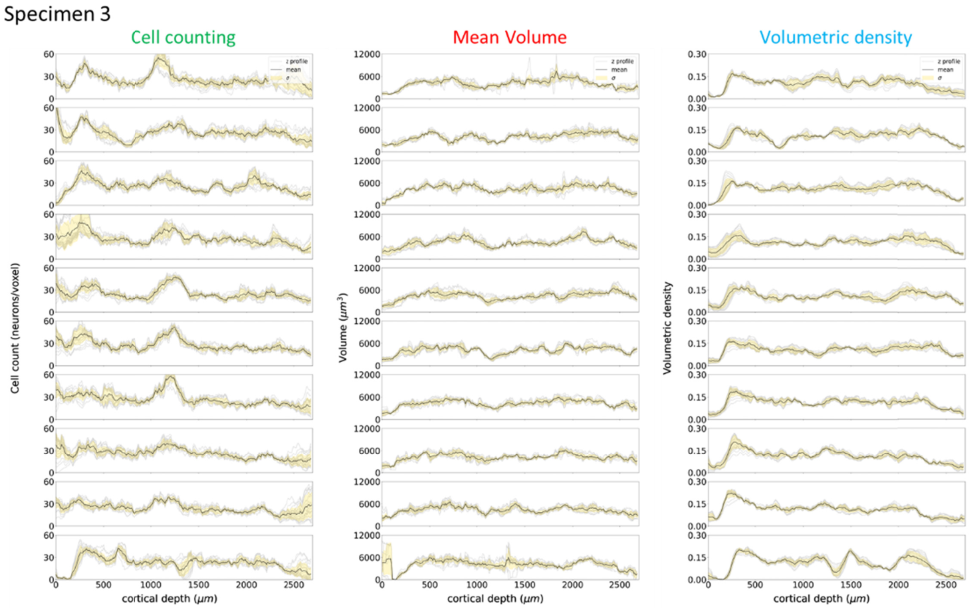


Fig. S11. Plots of cortical depth profiles (i.e. along Z) are shown together with the mean value and the one standard deviation for each of the 10 lines of specimen 3 showed in Fig S7

Manual segmentation of the grey matter. In order to identify the different layers of the grey matter of the cortex, we manually segmented the cell and mean volume distribution maps using the software Fiji (Supplementary figure S12). Big blood vessels, tissue holes/breakages, and imaging artifacts were not reckoned drawing the masks. Sample 4 shows a disruption of the structural organization of the cortex, making layer classification impossible, only grey matter were segmented. The total volume for each mask was obtained performing this operation for all the z-plane of the sample.

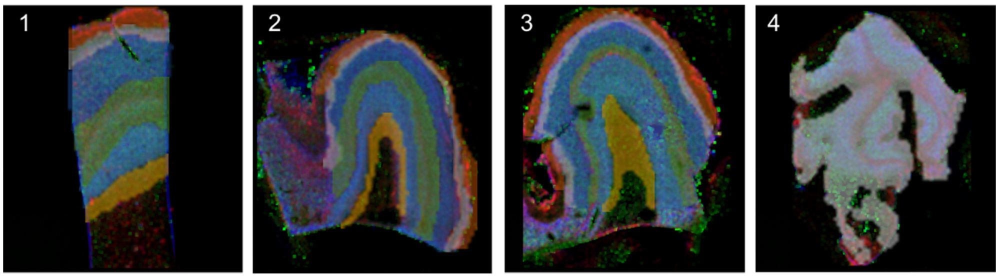


Fig. S12. Layer masks (L1 red, L2 white, L3 blue, L4 green, L5 light blue, L6 yellow) of the middle plane of samples 1,2, and 3. Grey matter mask **(**in grey) is shown for sample 4.